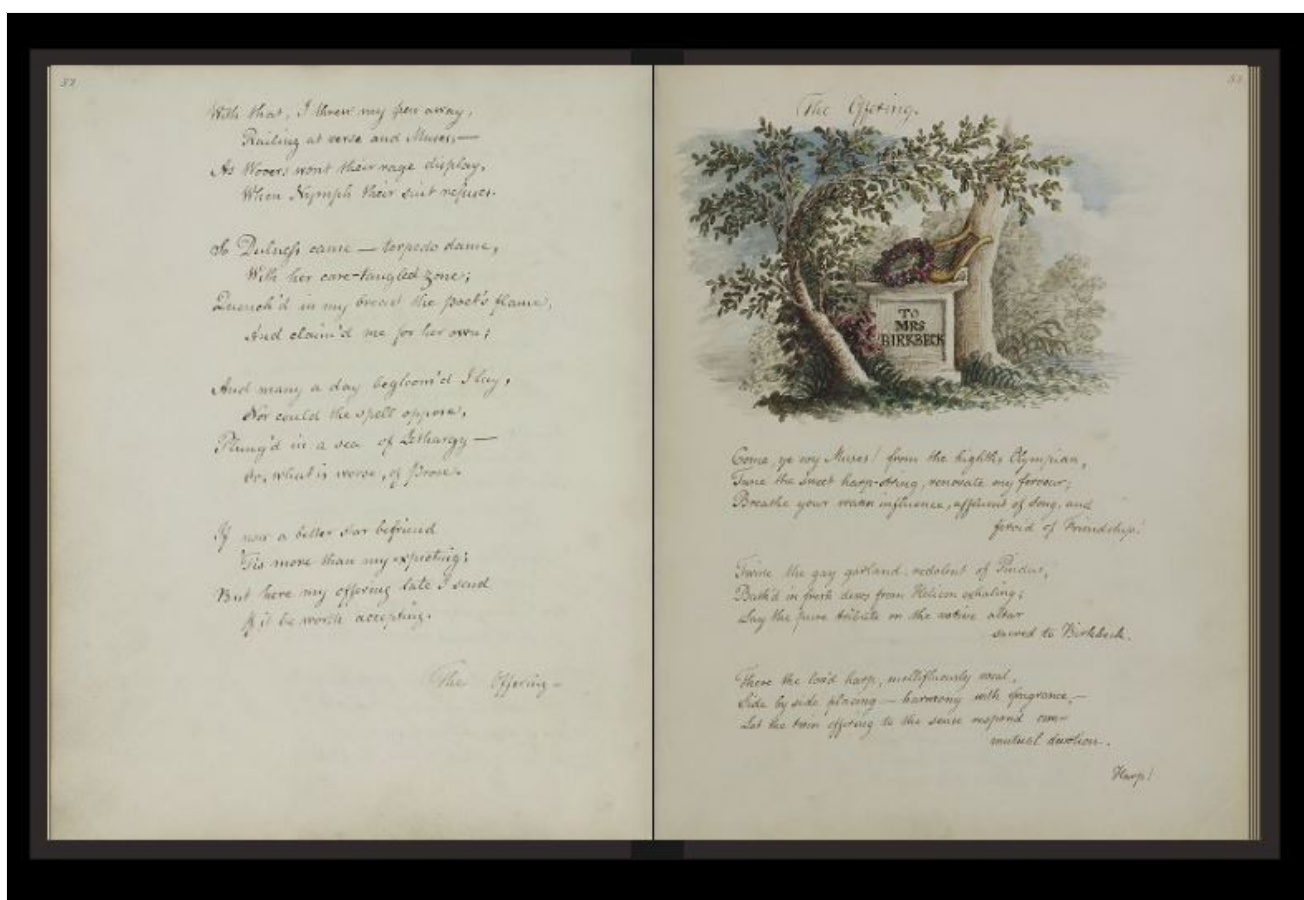


Open Cultural Data: Discussing Digitisation

This post was contributed by symposium organizers PhD candidate Hannah Barton, [Dr Joel McKim](#) and [Professor Martin Eve](#). The Open Cultural Data Symposium took place at Birkbeck on the 25 November 2016 and was co-sponsored by the [Vasari Research Centre for Art and Technology](#) and the [Birkbeck Centre for Technology and Publishing](#).

Birkbeck's recent Open Cultural Data Symposium was an opportunity to reflect upon several decades of major digitisation initiatives within UK cultural institutions. Academics, curators, archivists and IP specialists gathered in the Keynes Library to discuss the successes, ambitions and challenges of recent open access projects in some of the UK's most prominent museums, libraries and broadcast institutions.



The College has digitised the diary of Anna Birkbeck, the wife of George Birkbeck who founded the College

Adoption Beyond Access

The theme discussed by the first panel of the day was 'Adoption Beyond Access'. Dr Rebecca Sinker (Tate), Dr Mia Ridge (British Library) and researcher and curator Natalie Kane each set out to question what, beyond publication alone, institutions can do – or indeed are doing – to facilitate the use of their digitally accessible archives, collections and cultural data.

Dr Rebecca Sinker began by delineating the issues of scale and scope faced by institutions wanting to provide digital access to collections and facilitating associated outreach. Rebecca highlighted the importance of institutions committing to comprehensive infrastructural change and sustained investment when undertaking digitisation initiatives to avoid ad-hoc forays into collections access. However, Rebecca noted that resource limitations oftentimes make this an unattainable approach. Further, since it can take significant effort to establish digitisation and publications systems alone, the

importance of facilitating audience engagements with the published collections risks going unrecognised.

Yet the online publication of collections does not guarantee the material will be accessed by widened audiences. Using Tate's *Archives & Access* project as a case in point, Rebecca demonstrated how offering a range of 'entry points' to digitised collections can support varying levels of participation: from the additional access afforded by large-scale digital publication, to the entrees supported by online learning resources (such as explanatory films and blogs), to the in-person facilitated engagements, which can support audiences with differing levels of familiarity or confidence with cultural collections. Digital affordances allow new and exceptional modes of access, but some audiences may need support as they gain confidence and awareness of cultural collections before they take up that offer. In offering outreach in conjunction with digital access a more comprehensive cultural repositioning of cultural collections may be achieved in the long-term. However, with limited resources in mind, and a growing understanding of the role of outreach in engendering participation, advocacy remains necessary, the message being: publication and outreach in conjunction make for accessible – or rather accessed – open cultural data sets.



Dr Mia Ridge (British Library)

Dr Mia Ridge's presentation followed. Mia suggested that we begin by problematising the notion of cultural data. She asked the room to firstly take into consideration the quality of any data set that may be made open – what errors might it contain? Is it viable as structured open data? – and secondly to take into account the historicity of the set itself and its context of production. Does it contain any degree of cultural bias? Would it impart any degree of cultural bias if it was made open? To elucidate this point Mia references the digitally accessible Proceedings of the Old Bailey, 1674-1913 '*A fully searchable edition of the largest body of texts detailing the lives of non-elite people ever published, containing 197,745 criminal trials held at London's central criminal court*' – which is an amazing resource – detailed and accessible, but also a necessarily limited one. Exposure to open access data sets poses a risk, insofar as cultural bias may be created by over or under representation in open cultural data collections. The lives of non-criminal Londoners 1674-1913 are not so easily accessed, for instance, which may effect how literature or historical accounts are researched, written and interpreted. Further, individual issues of data set quality have the potential to impact on intra-institutional structured cultural data sets. "Every institution catalogues its archives in very different ways", noted

Mia, which will inhabit the ability for data sets to be joined up, and stymie the ambitions of those who wish to make horizontal journeys. She suggests that staff involved in open cultural data projects would benefit from increased understanding from scholars and other institutions alike – joined up conversations help to navigating this complex and dynamic topic, and events, such as hackathons and roadshows, can help in this regard as well as break down barriers to participation. Data in all forms, from published to collections to outcomes of practice sharing, flows both ways, Natalie Kane gave the final presentation of this panel; a fascinating talk that asked the room to challenge the politics of the archive, create parallel narratives, disrupt the space work occupies, interrogate categorisation and explore absence. “What might a postcolonial or feminist search engine look like?”, Natalie enquired. Pursuing this line of thinking, she showcased work from a range of artists who have explored this idea: 3D printing is mooted as a form of cultural reconstruction; a bust of Nefertiti is subject to a guerrilla-style digital scan as a challenge to colonial art theft; archival imagery is repurposed in unexpected ways, exploring absence and the tolerances in historical narratives. Natalie draws the audiences’ attention to Cécile B. Evans’ *Agnes*, a digital commission produced for the Serpentine Gallery’s website. *Agnes* is a bot in possession of an ‘aim-to-please’ character that playfully offers website visitors information both direct and tangential in nature. *Agnes*’ contributions can delight, confuse or frustrate and ultimately showcases disruption and frustrated forays into cultural collections. Natalie seizes upon this lack of structural totality as a distinguishing characteristic for anyone person exploring immaterial collections, and expounds the limits, but also the potential, such terms of distinction offer.

Legalities and Logistics of Digitisation



Fred Saunderson (National Library of Scotland), Bernard Horrocks (Tate) and Mahendra Mahey (British Library)

The second panel of the day focused on the “legalities and logistics” of implementing and maintaining large scale digitisation projects. Our three presenters, Fred Saunderson (IP Specialist at the National Library of Scotland), Bernard Horrocks (IP Manager at Tate) and Mahendra Mahey (Project Manager at the British Library Labs) outlined some of the pragmatic difficulties that can potentially stand in the way of a project’s lofty open access ideals. All three presenters dispelled the optimistic notion that the online environment could somehow alleviate the need for material spaces and physical “leg work” in relation to these projects. Fred Saunderson opened the panel and helped extend our discussion beyond

the confines of London. He highlighted the efforts made by the National Library to provide access to its collections to users across Scotland, despite being physically centred in Edinburgh. Online resources are not the only answer to this problem, he revealed, as onsite copyright licences can be considerably less restrictive and not all users gravitate to the digital realm. In response to these factors, the library has just opened a new film archive access centre at Kelvin Hall in Glasgow, with dedicated onsite terminals. While the library has currently been focusing on “low-hanging fruit” (material readily available for digitisation under various existing copyright exceptions, such as preservation requirements), Fred noted that there are considerable “scaling up” challenges ahead as the institution is committed to having a third of its collection available in digital form by 2025.

Bernard Horrocks focused on Tate’s recent Archives and Access digitisation project funded by the Heritage Lottery and involving approximately 53,000 archival items. While these items are all wholly owned by Tate, their copyright is not – a situation which introduces some considerable IP challenges. The scale of the problem was made clear when Bernard revealed that, despite belonging to 53 distinct collections, the items involved in the project could be traced back to some 1,500 rights holders. The number of human hours and amount of chasing involved in securing these rights (including a flight to Zurich) was clear and rather daunting, yet Bernard highlighted the level of success Tate achieved, with 98% of rights holders agreeing to some form of creative commons licences. Bernard emphasized the mix of due diligence, risk assessment and judicious use of copyright exceptions necessary for a project of this magnitude.

Finally, Mahendra Mahey outlined the impressive number of projects that have been supported by the British Library Labs since its inception. The BL Labs is an initiative funded by the Andrew W. Mellon Foundation and charged with encouraging public use of the library’s digital collections and data. The nature of the projects supported by the Labs varies considerably and Mahendra introduced a number of recent competitions, residencies, collaborations and events. Again, the success of these digital initiatives required considerable “real world” leg work, as raising awareness of the BL Labs was dependent on going out and talking to people. Mahendra emphasized the importance of “learning the story of the collection” as the origins and background history of the data in question largely determines the challenges involved in making it open.

Ethics and Organisation

The final panel of the day took a turn towards the ethical and organisational challenges surrounding open cultural data. Initially, we were supposed to be joined by a representative from HEFCE, who was sadly laid up with an illness. In his stead, however, Mia Ridge rejoined the panel, which also consisted of Dr Mark Coté (Lecturer in Digital Cultures, King’s College), and Bill Thompson, Head of Partnership Development, Archive Development, at the BBC.



3. Bill Thompson (BBC) and Mark Coté (King's College)

The paper given by Dr Coté was provocative. Arguing that many corporations are already collecting quantified behavioural data about users, he suggested that it was necessary for us to consider the opening of personal data as a site of political struggle. The suggestion seemed to be that because these corporations already act in this way, they remain the only entities who benefit from data analytics, leaving other actors out in the cold. But this suggestion came with many privacy challenges that left me feeling uncomfortable. I also was unclear over what political transformation we might see; do social justice organisations, for example, have the wherewithal and technical expertise to efficiently mobilise such data profiling – and how would it be used anyway?

Bill Thompson followed this with a talk about the institutional difficulties of working within an organisation such as the BBC at this time. Noting that the most recent charter for the organisation specifies little other than “programme making”, in contradiction to its founding remit of developing technologies for the public benefit, Bill pointed to the precariousness of his situation, working with the BBC archive; an amazing and diverse body of materials that are of enormous cultural significance. The day closed with discussions evolving into wine but one final point struck me, that Mia brought home. In this final twist on “data produced by humans as cultural data”, Mia noted that the temporal distance between recording and exposure is now so limited as to cause problems. In a previous era, if one wrote a personal diary, one would expect this to remain private. Not so of the public documentation of lives on social media, which can affect employment and many other aspects of one’s life. Indeed, though, how can we know which elements of our practices might be troublesome? How can we possibly evaluate the transactional benefit against the (only moderately) deferred risk? How does such open cultural data lead to a change in our own behaviours? These are the challenges of open cultural data that arose in the final panel.

More photos of the event are available on [flickr](#).

Further information:

- Professor Martin Eve
- Dr Joel McKim
- Department of English and Humanities
- Department of Film, Media and Cultural Studies
- Open Library of the Humanities