Routledge
Taylor & Francis Group

Cultural analytics in the UK: events data potential for the creative and cultural industries

Suzanne R. Black [a], Rosa Filgueira [b], Lesley McAra[a], Brendan Miles[c], Mark Parsons [a] and Melissa Terras [a]

aUniversity of Edinburgh, Edinburgh, UK; bUniversity of St Andrews, St Andrews, UK; cData Thistle, Edinburgh, UK

ABSTRACT

This article investigates the potential for novel research utilising data generated by the Creative and Cultural Industries (CCI) in the UK, focussing on the long tail of metadata associated with the UK's rich cultural events landscape. We conducted semi-structured interviews with 29 researchers and related domain experts to ascertain: (1) How cultural data is valued; (2) The landscape of cultural data; (3) How UK research can make better use of cultural events data; (4) The benefits and pitfalls of an evidence-based approach to cultural policy; and (5) The repercussions of the Covid-19 pandemic on how data-led work is positioned within the CCI. We advocate for the potential value of cultural events data to academic research, policy and industry, and also for a humanities-led approach. We suggest that a centralised cultural events data service for use in research, industry and policy is one way of supporting this.

Introduction

The Cultural and Creative Industries (CCI), as a major part of the UK economy, have attracted much research aiming to understand their structure, outputs and impact, from both economic and social perspectives. These industries generate a huge amount of data about events, venues, tickets and audiences and these data are extremely valuable to the field of cultural analytics, which incorporates aspects of "data science, the humanities, and media theory" for the "observation and analysis of global culture" (Manovich, 2020, p. 54). However, there is a debate about the value of cultural data to academic research, research for policy and research undertaken by CCI organisations with a push for increasingly data-led strategies met by critique of such positivistic and quantitative approaches.

This article intervenes in this debate by assessing the use and value of the long tail of metadata associated with the UK's rich cultural events landscape – theatre productions, music and comedy gigs, sporting fixtures, days out and more – as a data resource for academic research, policymakers and CCI organisations. It positions cultural events data as a

---

concrete example of the datafication of culture and asks how these data may be used in a way that enhances understanding of the CCI. To identify how these data are being used we undertook a qualitative study to ascertain the current landscape of data use by academic researchers, policymakers and CCI organisations, and canvassed opinions about its future direction, and the potential held for events data-led research.

We conducted semi-structured interviews with 29 individuals, including researchers working with – or interested in working with – large-scale data, as well as experts in data infrastructure, policy, the CCI, and the Arts and Heritage sectors. Using reflexive thematic analysis we found perspectives on: (1) How cultural data is valued by academic, social and industry research in the UK and how this relates to how culture is valued; (2) How large-scale cultural events data fit into the existing landscape of cultural data; (3) How UK research can make better use of cultural events data (skills and infrastructure); (4) The benefits and pitfalls of an evidence-based approach to cultural policy; and (5) The repercussions of the Covid-19 pandemic on how data-led work is positioned within the CCI. Our findings suggest that data-led research for the CCI is here to stay and that, in order to ensure that such research is as effective as possible, we need to keep sight of the critiques and interventions humanities approaches can offer, which will require investment in skills and infrastructure.

Humanities data research in the UK

Towards the datafication of culture

The UK has prolific Creative and Cultural Industries (CCI), defined as "those industries which have their origin in individual creativity, skill and talent and which have a potential for wealth and job creation through the generation and exploitation of intellectual property" (Department for Digital, Culture, Media & Sport, 2001). In recent years there has been a push towards datafication and this has also affected how culture is defined, measured and valued. Valtysson describes a situation in which a process of "[d]igitisation, digitalisation, datafication, platformisation" has occurred, starting with digitising data and resulting in "the increasing quantification of cultural content" followed by "the influence of major platforms as gatekeepers of cultural production and consumption" (2022, p. 786). These processes have influenced how data is used by the CCI, how policies are set by the government, and how data is understood as a resource.

Analysis of data practices within museums and galleries has found that the demands of the Covid-19 pandemic prompted a change in the "perception, use and importance of data within museums, and forced a recognition that strategic foresight and digital preparedness has significantly informed institutional ability to 'pivot' to digital delivery" (Noehrer et al., 2021, p. 5), leading to "a need for further research into specific measures which can assess the performance of digital products, to improve delivery, tailor content design and engagement practices around user experience, and to document success to satisfy and attract funders" (Noehrer et al., 2021, p. 9). This adoption of digital strategies by Arts and Heritage institutions mirrors the UK government's emphasis on increasing data use in the CCI, where:

> Collecting data not only allows cultural organisations to understand their audiences better but the data collected – audience demographic data, commercial transactional data,

metadata, crowdsourced data, locations-based data, and machine-generated data – can all be used to help develop business plans and organisation strategies. (DCMS, 2018; as cited in Wright & Gray, 2022, p. 805)

A similar approach to data use by the CCI can be seen in Nesta's argument that "[i]t is high time for a step-change in the approach of arts and cultural bodies to data and for them to take up and build on the management of so-called 'big data' in other sectors" (Lilley & Moore, 2013, p. 3). This trend of CCIs, policy-makers and commentators promoting big data over the last decade has led to the datafication and platformisation described by Valtysson, resulting in a "*Platform Society*" (Wright & Gray, 2022, p. 807, emphasis in original) where cultural participation is embedded within digital infrastructures so that "[d]istribution, circulation, and other work performed by cultural intermediaries are among the central concerns when it comes to social science research on cultural production" (Siciliano, 2022, p. 889). This means that the relationships of audiences to cultural objects are not mediated by institutions but, rather, "Artificial Intelligence (AI) technologies that involve machine learning algorithms replacing human judgment: algorithms as mediators of cultural participation undermine the monopoly of the traditional intermediaries and gatekeepers in the cultural field such as galleries, museums and specialist presses" (Rindzevičiūtė, 2022, p. 829). The result, writes Valtysson, is "*culture becoming digital*" (2022, p. 786, emphasis in original).

Positioned against these moves to datify the CCI are concerns around the implications of "evidence-based policy making" relying on the measurable impact of arts (Belfiore & Bennett, 2007, p. 135). Belfiore and Bennett argue that this approach relies on some assumptions, including the claim that positive impacts of the arts can be quantified with the result that "[p]ublic debate about the value of the arts thus comes to be dominated by what might best be termed the cult of the measurable" (2007, p. 137). Echoes of this "cult of the measurable" can be seen in cautions against ceding data collection, control and surveillance in a Platform Society where "it seems neutral and agnostic, but its architecture carries a particular set of ideological values" (Wright & Gray, 2022, p. 807). Valtysson similarly warns that the "tech giants that own and operate [platforms] are implemental in shaping the imaginaries of the future" (2022, p. 787) so that cultural policy needs to consider "privacy, transparency, individual and democratic control over data, quality and accessibility, equality, curation, inclusiveness, fair treatment, cultural rights, individual rights of data subjects, digital labour and working conditions and data discrimination" (2022, p. 791). These questions regarding the use of data in the CCIs and the wider datafication of culture lead to conversations about the place of data in our societies.

Events data

Data about cultural events offer an example of how the available data has changed along with the work that is done with that data and the repercussions this has for data-led analysis. Cultural events data can include metadata (about events taking place), locations-based data (where events take place), audience demographic data (who attended), crowdsourced data (how they felt about it) and commercial transactional data (attendee spend). Additionally, cultural events are mediated by platforms where attendees find out about events, book tickets and read reviews, and therefore participate in what Valtysson describes as "the increasing quantification of cultural content" and "the influence of major

platforms as gatekeepers of cultural production and consumption" (2022, p. 786). Cultural events data have been seized upon as vital by Manovich for use in cultural analytics where he defines events as "cultural happenings that have duration in time and involve multiple people: a music performance, an exhibition opening, a fashion show, a workshop, a weekend urban festival, a demonstration by a master coffee maker" (2020, p. 75), stressing their importance and their ubiquitous nature. Events data have unexplored value for "making fully visible the 'long tail' of cultural life – and placing on … culture maps cities, countries, groups, individual creators, and artifacts that have been left out from both contemporary and historical cultural narratives" (Manovich, 2020, p. 9). They also offer insight into how data is used and conceptualised in academia, industry and for policy: therefore also affecting future funding allocation and resourcing.

Sources of data already prevalent in cultural research include large-scale social datasets like the *Taking Part Survey*, the long-running DCMS survey of cultural participation in England (Department for Digital, Culture, Media & Sport, 2021) and *Active Lives*, Sport England's survey of sport and physical activity across England (Sport England, n.d.). Additionally, ticket sales and audience demographic data sourced from the Audience Agency, a non-profit organisation that supports the cultural sector in England and Wales by collating, comparing and contextualising data on cultural audiences ('Technical Glossary', 2019), have been used extensively and for multiple purposes, for example, alongside survey data to analyse customer booking patterns (Price et al., 2019), quantifying the impact of Covid-19 on the comic book industry (Fortnum et al., 2020), and on the cultural sector more broadly (Walmsley et al., 2022). The use of ticketing data (from the 2010s) is evidence of what Hanquinet et al. describe as the value of "digital by-product data" where "the information held by commercial organisations about consumer behaviours, particularly around consumption practices, seems to offer much more detailed, fine-grained, information on the social world compared with surveys and face-to-face interviews" (2019, p. 199).

Cultural events data have to date been put to use to understand issues of logistics in the Festival Mobility project, which combines events and transport data to better understand arts festival-related traffic and congestion (Ryan-Saha, 2020a; Ryan-Saha, 2020b). With the changing priorities wrought by the Covid-19 pandemic, this project pivoted to focus on public health and to organise festivals safely (Ali-Knight, 2021). Using a similar approach to investigate issues of inequality, the Culture and Communities Mapping project (Currie & Correa, 2021, 2022), combines multiple sources of event and social data to ask questions about social diversity in event provision. These projects use UK cultural events data, among other sources, from Data Thistle, a listings technology business that provides live events data (Data Thistle, n.d.).

In addition to taking the products of the CCI as research objects, the industries themselves generate data that is used in sociological research to characterise the CCI workforce (Beirne et al., 2017; Williams et al., 2022), to find ways of estimating cultural value (Comunian & England, 2020; Crossick & Kaszynska, 2016; McAndrew et al., 2020), and to expose cultural inequalities (Shaw, 2019, 2020). The issue of structural inequality in participation in the Arts and Heritage sectors has been a crucial one, quantified and explored extensively in *Culture Is Bad for You* by Brook et al. (2020).

Within academic research, Arts, Humanities and Social Sciences researchers are engaging with this "data-driven paradigm" (Schäfer & van Es, 2017, p. 11) while the CCIs

themselves (Noehrer et al., 2021) and research for policy purposes, as in the Creative Industries Policy and Evidence Centre (https://pec.ac.uk/), are also prioritising data-driven approaches involving both the use of large-scale cultural data and data-led methods for studying culture.

However, data use can exceed existing regulatory frameworks to, in effect, become "agents of implicit cultural policy, whereby the term 'policy' is used in its broad meaning to refer to the regulation and operationalisation of cultural practices underpinned by a specific set of values" (Bandinelli, 2022, p. 916). Hylland emphasises that it is often "major platform companies and tech companies", which "are supranational and with increasing power and influence at all levels" that are effectively shaping these implicit policies and makes a case for "scholarly work dedicated to exploring the entanglement between digital technologies, culture, media, and public policymaking" (2022, p. 815). Additionally, work has questioned the efficacy of data-driven approaches to cultural policy, such as Oman and Taylor's (2018) critique of non-rigorous approaches, which are used for selective advocacy although there are questionable claims for authority, with resulting problematic real-world impacts upon policy-making and funding decisions. Oman (2021) questions the use of health and well-being data, the assumptions underpinning their use (including presumed lack of bias), the limits of their application including lack of neutrality and objectivity, and potential misuse "as evidence in forms in which important decisions are made" (p. 5). In particular, Oman contextualises the overlaps between well-being data and cultural policy, showing how empirical accounts of uses of culture can be co-opted for both honourable and dishonourable means (pp. 229-263). In the humanities more widely, Hall interrogates the move from ideology and critique toward more positivistic, quantitative, and empirical modes of analysis, but stresses we need to develop new forms of criticism to engage fully with this digital shift and understand its ramifications (2013).

Against this global entanglement of data and culture, Belfiore and Bennett see the problem of "the extent to which arts organisations and funding bodies have allegedly subordinated (at least at the level of rhetoric if not practice) purely artistic considerations to a preoccupation with art's social or economic impacts" (2007, p. 147) and advise that "the humanities can make a significant contribution to the construction of a strong and coherent theoretical framework for the elaboration of more rigorous arts impacts evaluation methods" (2007, p. 148). We position this study in the context of the datafication of culture and questioning of "the cult of the measurable" (Belfiore & Bennett, 2007, p. 137) as an intervention into the ways in which cultural events data may be useful from a humanities research perspective.

Research design and method

Interviews

Our interviews were designed to ascertain the ways in which researchers and those who work with data across academia, policy and industry engage with cultural data, the types of questions they try to answer, and any barriers encountered. To this end, we interviewed researchers already working with large-scale data, those who would like to, and experts in data infrastructure, policy, the CCI, and the Arts and Heritage sectors.

Author One conducted semi-structured interviews with 29 participants and asked questions from a prepared interview schedule. The semi-structured nature of the interviews allowed us to tailor the questions to the participants' expertise. For example, we asked academic researchers if and how they currently use events data or other data relating to the CCI in their research, and how they would prefer to access and retrieve large-scale datasets. For participants who work in the area of data infrastructure we focussed on the infrastructural requirements of storing and making available large-scale datasets.

Interviews were conducted over Zoom between Wednesday 2 March and Thursday 5 May 2022. The audio and video of the interviews were recorded and then transcribed verbatim by the Project Assistant. In total, 25 h of interview material was transcribed, giving some 158,000 words of content to analyse via reflexive thematic analysis (Braun & Clarke, 2021).

We also undertook five unstructured interviews with major industry events providers. Interviews were conducted over Zoom or Microsoft Teams by Authors Four and Six during June 2022. Due to the commercial sensitivity of these discussions, the interviews were not recorded but extensive notes were taken by both interviewers, and compared and contrasted immediately afterwards.

Ethics issues and approval

Ethics approval was obtained from the School of Languages, Literatures and Cultures Research Ethics Committee, University of Edinburgh, approval number LLC0000TERRA-S00AH/W007533/1. In advance of each interview we sent an information sheet to each participant and received a signed consent form. Interview transcripts were returned to participants who were given time to make comments or redactions. Given humanities data research is a small field, care has been taken here to not reveal participants' identities by avoiding direct links between quotations and published work.

For the interviews with industry events providers, we sent study information in advance to each participant and received confirmation via email that they were willing to proceed. We confirmed that we would only report back thematically, after coding the combined notes from the interviews.

Participants overview

We used the snowball method of recruitment (Biernacki & Waldorf, 1981) by approaching potential participants via contacts and asking them to suggest other participants until new names ceased to be suggested. We engaged participants who work with large-scale cultural data as well as those with related subject expertise. The participants had backgrounds in academic research, the GLAM and Heritage sectors, economics, and roles at organisations that champion the implementation of data and data science. The academic researchers (14 out of the 29 participants) work in audience research, the creative and cultural economies, cultural equality, cultural policy, the sociology of culture and culture generally. The participants who were not fully engaged in academic research brought expertise in data science, software engineering, technological infrastructure, designing collaborative data projects, the generation and analysis of cultural data, data curation, and influencing CCI policy.

Interview analysis

Data from the interviews were analysed using reflexive thematic analysis. Author One took a semantic approach "where the analysis explores meaning at the more surface, explicit, or manifest level" (Braun & Clarke, 2021, p. 10) combined with an inductive approach "where the analysis is located within, and coding and theme development are driven by, the data content" (p. 10), reflecting the perspectives of the participants. This was done within a realist framework "where analysis aims to capture truth and reality, as expressed within the dataset" (p. 10). This variation of reflexive thematic analysis – semantic, inductive and realist – was chosen since the aim of this research is to canvas for expert opinions from the participants about their own knowledge domains. Through the analysis we sought to identify commonalities across the data regarding how those working with cultural data across academia, policy and the CCI understand the current state of data-led practices and how they can be improved. We acknowledge that the perspective of Author One as an academic researcher may influence the data gathered from participants who work across academia but also industry and policy.

Analysis was guided by the six phases of Braun and Clarke's (2021, p. 35) reflexive thematic analysis:

- Familiarising yourself with the dataset: listening to the interview audio and reading the transcripts.
- Coding: breaking the interview data into semantically coherent units and assigning each unit a short code; rereading and revising the material in each code for consistency.
- Generating initial themes: identifying data across the codes that showed "patterns of shared meaning underpinned by a central organising concept" (Braun & Clarke, 2019, p. 589).
- Developing and reviewing themes: returning to the initial themes to check for coherence, connecting the themes with each other and with contextual literature; establishing the final five themes.
- Refining, defining and naming themes: developing informative names for each theme and choosing illustrative quotes.
- Writing up: constructing narratives to reflect the content of the interviews and their relation to existing literature while taking care to avoid summarising topics.

The resulting five themes are described below.

Findings and discussion

In the analysis of the interview data we identified five themes: (1) How cultural data is valued by academic, social and industry research in the UK and how this relates to how culture is valued; (2) How large-scale cultural events data fit into the existing landscape of cultural data; (3) How UK research can make better use of cultural events data (skills and infrastructure); (4) The benefits and pitfalls of an evidence-based approach to cultural policy; and (5) The repercussions of the Covid-19 pandemic on how data-led work is positioned within the CCI.

The values of cultural data

This theme highlights the ways in which cultural events data are valuable for a range of purposes, and what this means for how culture itself is valued and evaluated.

A sense of the value of cultural data – and events data specifically – was frequently evident, not only to academic researchers but to policymakers and representatives of the CCI themselves. Participants spoke about the potential benefits for policy-makers where "in policy terms, there's a lot of interest in creative clusters … events data could be the beginning of a characterisation of the cultural make-up of the locality, which gives them a richer description of what these clusters are" (P08), for local authorities "who are also looking at their events strategy [and] would maybe want to see what the landscape is like … where they could find a niche" (P23), and in "making policy decisions based on historical gaps in provision, for example, funding gaps" (P16), finding uses for events data in evaluating, planning and decision-making.

This positive attitude to cultural events data is expressed in the following quote by Participant 15:

> I've got this, I suppose, utopian vision of live events affecting just about everything that anybody ever does anywhere. Now clearly that's an exaggeration. But actually, it really does affect so many people in different ways and so many different industries. And getting better at being able to provide that information, to learn from that information, to link, to hook it up to different datasets, it just shows an incredible world of possibilities that we're only just scratching the surface of. (P15)

In the above quotes, the participants make a case for the value of access to cultural events data for policy-makers and the CCI themselves, especially the Heritage, Arts, and tourism sectors. The usefulness of cultural events data is also reflected in scholarship on the value to events providers of understanding "the relationship between intention, ticket purchase and attendance" across multiple artforms and demographics (Price et al., 2019, p. 222).

As well as speaking to the value of events data, Participant 07 introduced the importance of data in valuing UK culture more widely:

> [O]ne of the key things is about the soft power of UK culture globally in terms of tourism and bringing in revenue and business. And so, if we had a data set that could help underpin claims about soft power … I think it would be really, really helpful and would stop us looking like we were just telling and let us show as well. (P07)

There is a sense that the participants value cultural events data as one element in their toolkit to enable better evidence-based research, decision-making and operation of the CCI.

A precursor to valuing cultural data and culture more widely is to define them. Several of the participants brought this up early in their interviews, asking "what is a cultural event?" (P13) and "what is a cultural amenity or what is a venue and what [do] you count?" (P02). While data definitions are a normal first step in data analysis, these questions play into a larger conversation about evaluating culture. Belfiore and Bennett, in their work on the social impact of the arts, link government funding of the arts to a focus on the measurable impact of arts in "evidence-based policy making" (2007, p. 135). The topic of cultural data as evidence recurs throughout the themes, although this requires careful consideration to use these data appropriately and holistically, stating its biases and understanding any limitations.

How cultural events data fit into the current data landscape

This theme concerns how large-scale cultural events data might fit into the existing landscape of cultural data. Specifically, it reflects a tension between cultural events data and other forms of transactional data – such as tickets sales – and the existing reliance on national survey data.

Participant 02 describes how access to comprehensive national events data would complement existing survey data like *Taking Part* and the *Scottish Household Survey* where "it would allow for a much more granular understanding of the cultural life really than a broad artform categorisation allows for" (P02), with Participant 17 stating that "it would help to connect up a lot of the disparate, case studies-style research that goes on" (P17). This reflects the findings of Hanquinet et al. who compared the usefulness of transactional data about ticketing with social survey data (from *Taking Part* and *Active Lives*) to find that "the information held by commercial organisations about consumer behaviours" (2019, p. 199) is "particularly useful for shedding light on activities that are hard to gain insight into from surveys" (2019, p. 214).

Another participant spoke more strongly about what cultural events and other CCI transactional data could offer:

> [T]hese new data sources say something new and more useful and more timely instead of waiting for years for a survey to come out. And a survey … if you're relying on random samples, you don't get to know about niches in great detail. (P10)

This combination of qualitative and quantitative data is further explained in the literature:

> One of the challenges of big data approaches is that the data collection occurs separately from and/or prior to the framing of the questions that it can answer. This type of combined analysis offers a means of formulating specific research questions that can be reapplied to the transactional dataset. Conversely, the volume of quantitative data allows for recognition of patterns which are not self-reported by interviewees. This can inform the framing and interpretation of the qualitative responses. (Price et al., 2019, pp. 233–234)

Participant 25 also finds that quantitative work can be used to problematise survey data:

> I'm really interested in the question of the extent to which data-based work tells you things that you don't want to hear. So I think what I like about working with these kind of data sets is that there's the possibility to have some sort of concrete data-led answer to a question you might have. (P25)

Whether using cultural events data to complement or question survey data, Participant 09 sums up the feeling that "there are things that aren't necessarily formal data sets, in the sense that the *Annual Population Survey* is a dataset, but it's certainly very relevant data to understand creative and cultural spaces" (P09). This theme reflects an appetite for new data sources that can be used to work with and against existing sources.

Making better use of cultural events data (skills and infrastructure)

This theme gathers the participants' concerns that there are skill and infrastructure gaps preventing the use of cultural events data and what could be achieved if these gaps were filled.

Participant 16 describes a situation in which the UK is world-leading in certain skills but has the potential to do much more:

[W]e really do need some cultural economists and more quantitative sociologists in the UK. There's a real gap. I mean there's a global gap clearly. We're probably world leading actually in this but even so, you can probably count them on one hand, the senior academics working this area. So there's a real pipeline issue and there's a real need for training but training isn't the only answer. (P16)

As a way to work around differing skills levels, Participant 28 describes a recent data analysis project where Jupyter notebooks were created by research software engineers for use by members of the team who were less proficient in computational analysis so that "they'll work in Python, but it will be using a template notebook that's already been set up" (P28). From this it can be inferred that researchers using large-scale cultural data approach it with varying levels of skill in working with data and from different disciplinary perspectives. While there is an ongoing aim to improve the data skills of Arts and Humanities researchers (Flanders & Jannidis, 2015; McGillivray et al., 2020; Zhang et al., 2021), Participant 04 cautions that:

Where the number of people who actually have the skill set to work with this kind of data to a realistic extent and who have any time is very limited. And it would be great to up-skill people. I would be delighted if this research community increased in size, but we need to be realistic about how long that's going to take. (P04)

From the participants there emerged the sense that infrastructural rather than individual efforts would greatly enhance access to and the ability to utilise cultural events data. Participant 05 admits that "[w]e have a really good infrastructure for certain things in the UK" but that this is only "provided you know what you're looking for" (P05). Participant 08 makes the point that "to build an infrastructure which allows researchers to access that data more easily … without this infrastructure, you've just been left to identify these opportunities on your own and negotiate them" (P08). This is echoed by Participant 18 who says that:

I think one of the things that we've been very slow to realise in the UK is data as an infrastructure, because if you don't have it, you can't do your science … we're moving to a point where data perhaps needs to be seen as an infrastructure element. (P18)

Beyond access, Participant 28 explains the benefits of sharing data resources where multiple teams enhancing the same dataset leads to a "cumulative benefit" from "the more time that's spent analysing the data" (P28). There was also mention of the value of combining data sources: "Where I think the particular benefits of this stuff would be around combining it with other data sources, particularly ONS [Office for National Statistics]" (P04).

As elaborated on in section 4.4, UK cultural policy is focused on "the technological infrastructure that is considered fundamental to the future prosperity of the UK" (Wright & Gray, 2022, p. 800) and this mirrors the global situation where "platformisation" (Valtysson, 2022, p. 786) has led to a situation where a handful of tech companies "serve as influential gateways to cultural knowledge" (2022, p. 787). Since "national cultural policy faces obvious challenges when key actors, like the major platform companies and tech companies, are supranational and with increasing power and influence at all levels" (Hylland, 2022, p. 815), UK academia requires an understanding of similar platforms and infrastructure.

Several of the participants were enthusiastic about a centralised data service as an infrastructural solution to increasing access to cultural events data. Speaking about the difficulties of locating UK publishing data, Participant 07 expressed frustrations that "a lot of it was about googling, and then trying to paper trail back to see who owned that information, who the data officer was, where it was, if it didn't exist, why? So there were gaps" (P07). They went on to suggest that "if we had a central repository for that kind of data, it would be so helpful because quite often the data is brilliant" (P07).

Participant 01 elaborated on both the potential benefits of greater access to events data as well as some of the logistical concerns:

> [T]here's so many events happening UK wide. So … how would we make it a live feed of data available? And how would we preserve this data? What format would it be kept in and how do we make sure that all the events around the country come together in a format that we can bring together as a single data set? (P01)

This theme highlights the skills and infrastructure gaps that exist around working with cultural events data from sociological and Arts and Humanities perspectives in UK academia and suggests that a centralised data resource and – as in Participant 28's quote – interdisciplinary modes of working can mitigate these gaps, but this requires intervention at a level higher than that of the individual researcher.

The benefits and pitfalls of an evidence-based approach to cultural policy

This theme offers insight into the wider repercussions of the quantification and datafication of culture, which is viewed by some of our participants as positive while others view it as part of a larger trend with potential pitfalls.

Participant 16 speaks to the value of cultural data to evidence-based policymaking where "it just provides an evidence base that is pretty objective, doesn't it, about what's happened where and what hasn't happened where? So I think from a policy perspective it's hugely important" (P16). Conversely, Participant 17 lists potential harms that could arise from using cultural events data where they might "be a threat to small venues that are not meeting … particular thresholds of attendance or success", or "if it got into the wrong hands, if city councils are trying to make decisions about cutting buildings" or "the potential for the data to be manipulated" to support "arguments about arts being elitist" (P17).

The move towards quantifying culture and the potential pitfalls of this have been noted with regards to the UK Government's cultural strategy where the Department for Culture, Media and Sport became the Department for Digital, Culture, Media and Sport in 2017 amidst the "increasing centrality of digital infrastructures and platforms for cultural policy and cultural policy research" (Wright & Gray, 2022, p. 799).[1] This "datafication" (Valtysson, 2022, p. 786) is alluded to by Participant 21 who questions whether a focus on measurability leads to losing sight of humanitarian consequences:

> I think people rather get obsessed with data and the ability to analyse it and I do worry sometimes that we spend far too much time on baselining and getting as many statistics as we possibly can when actually … it's far more to do with intervention and monitoring, evaluating the facts of those interventions. (P21)

Here, Participant 21 echoes the warning of Belfiore and Bennett about an increasingly evidence-based approach where "it is those disciplines primarily concerned with

measurement, namely, economics and statistics, which are looked upon to find the evidence that will finally prove why the arts are so important to individuals and societies" with the result that "the humanities are of little use in this investigation" (2007, p. 137). If the use of cultural events data is going to be more widespread, as the participants in section 4.3 suggest is the case, then it is imperative that a humanities perspective is included, critiquing the available data, its coverage, the methods used to analyse it, and the ramifications of the use of any results, to understand the implications this type of events data analysis has for the cultural sector.

The effects of the covid-19 pandemic on data-led work in the CCI

This theme reflects the specific historical moment of the early 2020s in which any research must respond to the changes wrought by the Covid-19 pandemic. We found that the participants reflected on the pandemic as a pivotal point to establish the importance of quantitative data and embedding data gathering and analysis practices into both the Arts and Heritage sectors and academic research.

Data work has been important for understanding the effects of the pandemic and planning support, recovery and the future directions of the CCI. Researchers turned to large-scale cultural data to ask questions around mitigating covid effects, like "[h]ow do we run an event in the time of covid?" (P03). The value of data work is described amidst the uncertainty of a post-covid landscape where "booking patterns went from six months on average in advance to six weeks, then down to six days" (P15). At the same time, there is a sense that the pandemic exposed some gaps in existing cultural data provision, with Participant 16 reflecting that understanding and providing financial support for the sector was hampered by a lack of existing workforce data around freelance populations with "[h]uge gaps especially around the number of freelancers working in different sub-sectors, and the hugely complex role that freelancers play in supporting the arts and cultural sector and wider creative industries" (P16).

This focus by the participants on the importance of cultural data for industry, research and policy reflects a situation in museums and galleries where there has been more emphasis on metrics and data-driven decisions where "COVID-19 has altered the perception, use and importance of data within museums, and forced a recognition that strategic foresight and digital preparedness has significantly informed institutional ability to 'pivot' to digital delivery" (Noehrer et al., 2021, p. 5). The Covid-19 pandemic has only accelerated the trends described in section 4.4 on the benefits and pitfalls of an evidence-based approach to cultural policy.

Speaking to another pivotal and contemporary issue, our industry contacts also expressed a concern for the climate crisis, and that improved information sharing, and the ability to compare and link their datasets with other information could improve sustainability. For example, multi-country and multi-venue tours could be planned in a way that best utilises sustainable travel, not only for artists, but for dedicated fans, too, which would reduce the environmental impact of large-scale events. Additionally, they raised concerns around the effect that more extreme weather will have on future events. There was an acknowledgement of the carbon footprint of the events industry, and a hope that data-led approaches could assist the industry in understanding how to be more efficient and sustainable in their practices.

Limitations and future work

This study has a number of limitations. In the research design, the participants interviewed were drawn from those already working in the space of cultural data or adjacent to it. Although we attempted to gauge the potential for expanding this field, we cannot accurately know how many other researchers, practitioners and policymakers are open to expanding their skills in this direction or forging interdisciplinary working groups. The majority of the participants hold roles in academic research, research for policy or similar, and fewer are policymakers or industry professionals. Future interrogations of this space would benefit from expanding the interview pool.

The results reported here are likely biased towards data-driven approaches given that the participants we interviewed are already working with cultural data. While we have tried to reflect the positive and negative views of the participants and include dissenting voices from the literature, future work in this area could consider seeking a wider range of perspectives, including any issues of data-biases, representation, or potential misuse of this type of analysis, which have been highlighted in other research areas that have utilised data analysis to drive policy and related decision-making regarding funding (Oman, 2021).

Finally, the suggestion by some of the participants that infrastructural investment into a central cultural data service would greatly benefit research, policy and industry regarding the CCI (particularly for Humanities researchers) suggests that further research is required into the practicalities of such infrastructure. This should include technical specifications, funding and access models, and the willingness of industry partners to supply the necessary data, given commercial, ethical and privacy sensitivities. In addition, fully understanding the training and support needs of Humanities researchers will be necessary if their perspectives and approaches are to be successfully brought into this data-led space.

Conclusions

In order to ascertain the present state of cultural analytics work in the UK, the scope for expanding upon this work, and areas of further concern, we interviewed 29 participants from across academia, heritage, data science and policy domains. From within the range of responses we have been able to discern the priorities of the UK's cultural analytics community as well as those of interested stakeholders.

We identified a variety of work in or adjacent to cultural analytics as well as challenges around the availability of data, skills gaps and methodological concerns. Despite these challenges there was a marked appetite for more work in this direction. Suggestions about how to mitigate these challenges included being wary of relying solely on positivistic and quantitative approaches, and understanding the value of humanities perspectives on data. One suggestion for how to do this was to invest in infrastructure to enable Arts and Humanities and Social Science researchers, as well as those working in the CCIs, to gain access to cultural events data in a way that reduces the technological burden upon them. Establishing a data infrastructure and support to better provide researchers, industry and policymakers with large-scale cultural data would enable the UK to make the most of its extraordinarily rich data output and continue to be a world-leader in cultural analytics, based on its world-leading cultural events industry and

economy. However, it is necessary to bear in mind the limitations of data-led approaches to culture and cultural value, to ensure that they are used appropriately in order to support research on, and the work of, our Creative Industries.

Note

1. It should be noted that in 2023 the department has reverted back to being the Department for Culture, Media and Sport, with the "Digital" under the purview of the Department for Science, Innovation and Technology. While this change signals a change of categorisation, and perhaps priorities, for the UK government, data-led practices are now so integrated into the CCI that we do not see this trend reversing in the short to medium term. It will be necessary for future research in this area to note the repercussions of this departmental name change.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Suzanne R. Black http://orcid.org/0000-0001-5479-0528
Rosa Filgueira http://orcid.org/0000-0002-5715-3046
Mark Parsons http://orcid.org/0000-0003-4097-7468
Melissa Terras http://orcid.org/0000-0001-6496-3197

References

Ali-Knight, J. (2021). *Modelling Mobility and Infection: The Reopening of Edinburgh's Festivals*. University of Edinburgh. https://publichealthscotland.scot/repository/modelling-mobility-and-infection-the-reopening-of-edinburghs-festivals/.
Bandinelli, C. (2022). Dating apps: Towards post-romantic love in digital societies. *International Journal of Cultural Policy*, *28*(7), 905–919. https://doi.org/10.1080/10286632.2022.2137157
Beirne, M., Jennings, M., & Knight, S. (2017). Autonomy and resilience in cultural work: Looking beyond the 'creative industries'. *Journal for Cultural Research*, *21*(2), 204–221. https://doi.org/10.1080/14797585.2016.1275311
Belfiore, E., & Bennett, O. (2007). Rethinking the social impacts of the arts. *International Journal of Cultural Policy*, *13*(2), 135–151. https://doi.org/10.1080/10286630701342741
Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, *10*(2), 141–163. https://doi.org/10.1177/004912418101000205

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Braun, V., & Clarke, V. (2021). *Thematic Analysis: A Practical Guide*. Sage.

Brook, O., O'Brien, D., & Taylor, M. (2020). *Culture is Bad For You: Inequality in the Cultural and Creative Industries*. Manchester University Press.

Comunian, R., & England, L. (2020). Creative and Cultural Work Without Filters: Covid-19 and Exposed Precarity in the Creative Economy. *Cultural Trends*, *29*(2), 112–128. https://doi.org/10.1080/09548963.2020.1770577

Crossick, G., & Kaszynska, P. (2016). *Understanding the Value of Arts & Culture: The AHRC Cultural Value Project*. Arts and Humanities Research Council. https://www.ukri.org/publications/ahrc-cultural-value-project-report/.

Currie, M., & Correa, M. M. (2021). *The Culture and Communities Mapping Project*. Palgrave Macmillan. https://link.springer.com/book/10.1007978-3-030-88651-6.

Currie, M., & Correa, M. M. (2022). Tangibles, intangibles and other tensions in the culture and communities mapping project. *Cultural Trends*, *31*(1), 88–106. https://doi.org/10.1080/09548963.2021.1910491

Data Thistle. (n.d.). *Events Data*. Retrieved October 28, 2022, from https://www.datathistle.com/about/events-data/.

Department for Digital, Culture, Media & Sport. (2001). *Creative Industries Mapping Documents 2001*. GOV.UK. https://www.gov.uk/government/publications/creative-industries-mapping-documents-2001.

Department for Digital, Culture, Media & Sport. (2021, June 24). *Taking Part Survey*. GOV.UK. https://www.gov.uk/guidance/taking-part-survey.

Flanders, J., & Jannidis, F. (2015). *Knowledge Organization and Data Modeling in the Humanities*. http://www.wwp.northeastern.edu/outreach/conference/kodm2012/flanders_jannidis_datamodeling.pdf.

Fortnum, J., Verhoeff, R., & Cowley, D. (2020). *UK Comics Creators Research Report*. The Audience Agency. https://hannahberry.co.uk/wp-content/uploads/2020/07/UK-Comics-Creators-research-report-2020.pdf.

Hall, G. (2013). Toward a postdigital humanities: Cultural analytics and the computational turn to data-driven scholarship. *American Literature*, *85*(4), 781–809. https://doi.org/10.1215/00029831-2367337

Hanquinet, L., O'Brien, D., & Taylor, M. (2019). The coming crisis of cultural engagement? Measurement, methods, and the nuances of niche activities. *Cultural Trends*, *28*(2-3), 198–219. https://doi.org/10.1080/09548963.2019.1617941

Hylland, O. M. (2022). Digital cultural policy. The story of a slow and reluctant revolution. *International Journal of Cultural Policy*, *28*(7), 813–828. https://doi.org/10.1080/10286632.2022.2137148

Lilley, A., & Moore, P. (2013). *Counting What Counts: What Big Data Can do for the Cultural Sector*. Nesta. https://www.nesta.org.uk/report/counting-what-counts-what-big-data-can-do-for-the-cultural-sector/.

Manovich, L. (2020). *Cultural Analytics*. MIT Press.

McAndrew, S., O'Brien, D., & Taylor, M. (2020). The values of culture? Social closure in the political identities, policy preferences, and social attitudes of cultural and creative workers. *The Sociological Review*, *68*(1), 33–54. https://doi.org/10.1177/0038026119871401

McGillivray, B., Alex, B., Ames, S., Armstrong, G., Beavan, D., & Ciula, A. (2020). *The Challenges and Prospects of the Intersection of Humanities and Data Science: A White Paper from The Alan Turing Institute*. The Alan Turing Institute. https://figshare.com/articles/online_resource/The_challenges_and_prospects_of_the_intersection_of_humanities_and_data_science_A_White_Paper_from_The_Alan_Turing_Institute/12732164.

Noehrer, L., Gilmore, A., Jay, C., & Yehudi, Y. (2021). The impact of COVID-19 on digital data practices in museums and art galleries in the UK and the US. *Humanities and Social Sciences Communications*, *8*(1), https://doi.org/10.1057/s41599-021-00921-8

Oman, S. (2021). *Understanding Well-Being Data: Improving Social and Cultural Policy, Practice and Research*. Springer Nature. https://doi.org/10.1007/978-3-030-72937-0

Oman, S., & Taylor, M. (2018). Subjective well-being in cultural advocacy: A politics of research between the market and the academy. *Journal of Cultural Economy*, *11*(3), 225–243. https://doi.org/10.1080/17530350.2018.1435422

Price, S., Perry, R., Mantell, O., Trinder, J., & Pitts, S. (2019). Spontaneity and planning in arts attendance: Insights from qualitative interviews and the Audience Finder database. *Cultural Trends*, *28*(2-3), 220–238. https://doi.org/10.1080/09548963.2019.1617943

Rindzevičiūtė, E. (2022). AI, a wicked problem for cultural policy? Pre-empting controversy and the crisis of cultural participation. *International Journal of Cultural Policy*, *28*(7), 829–844. https://doi.org/10.1080/10286632.2022.2137151

Ryan-Saha, J. (2020a, February 17). City Mobility – Decisions Based on Events Data. *Data Thistle*. https://datathistle.medium.com/city-mobility-decisions-based-on-events-data-1a8d5fe51e09.

Ryan-Saha, J. (2020b, April 20). *Festival Mobility: Decisions with Data*. The University of Edinburgh. https://www.ed.ac.uk/edinburgh-international-data-facility/updates-events/festival-mobility-decisions-with-data.

Schäfer, M. T., & van Es, K. (2017). *The Datafied Society: Studying Culture Through Data*. Amsterdam University Press. https://www.jstor.org/stable/j.ctt1v2xsqn.

Shaw, K. (2019). Writing the Northern Powerhouse: Evaluating the Northern Writers' Awards as a potential model of intervention for addressing regional representation in the 'London-Centric' UK literary industry. *Creative Industries Journal*, *12*(1), 3–13. https://doi.org/10.1080/17510694.2018.1523667

Shaw, K. (2020). Common people: Breaking the class ceiling in UK publishing. *Creative Industries Journal*, *13*(3), 214–227. https://doi.org/10.1080/17510694.2019.1707521

Siciliano, M. L. (2022). How gatekeeping became digital: Infrastructural barriers to participation in conventional and platformized cultural production. *International Journal of Cultural Policy*, *28*(7), 888–904. https://doi.org/10.1080/10286632.2022.2137154

Sport England. (n.d.). *Active Lives*. Retrieved October 28, 2022, from https://www.sportengland.org/research-and-data/data/active-lives.

Technical Glossary. (2019). Technical glossary. *Cultural Trends*, *28*(2-3), 263–264. https://doi.org/10.1080/09548963.2019.1634322

Valtysson, B. (2022). The platformisation of culture: Challenges to cultural policy. *International Journal of Cultural Policy*, *28*(7), 786–798. https://doi.org/10.1080/10286632.2022.2137150

Walmsley, B., Gilmore, A., O'Brien, D., & Torreggiani, A. (2022). *Culture in Crisis: Impacts of Covid-19 on the UK Cultural Sector and Where We Go From Here*. Centre for Cultural Value. https://www.culturehive.co.uk/wp-content/uploads/2022/01/Exec_Summary_Culture_in_Crisis.pdf.

Williams, K., Lester, L., & Seivwright, A. (2022). The invisible architecture of creative and cultural work: The relationship between miscategorisation and sector wellbeing during COVID-19. *Creative Industries Journal*, 1–19. https://doi.org/10.1080/17510694.2022.2106091

Wright, D., & Gray, C. (2022). Culture is digital and the shifting terrain of UK cultural policy. *International Journal of Cultural Policy*, *28*(7), 799–812. https://doi.org/10.1080/10286632.2022.2137149

Zhang, Y., Su, F., & Hubschman, B. (2021). A content analysis of job advertisements for digital humanities-related positions in academic libraries. *The Journal of Academic Librarianship*, *47*(1), https://doi.org/10.1016/j.acalib.2020.102275