



Doyle, G. (2018) Television and the development of the data economy: data analysis, power and the public interest. *International Journal of Digital Television*, 9(1), pp. 53-68.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/156499/>

Deposited on: 1 February 2018

Enlighten – Research publications by members of the University of Glasgow_
<http://eprints.gla.ac.uk>

**Television and the Development of the Data Economy:
Data analysis, Power and the Public Interest**

Gillian Doyle

January 2018

Abstract:

Data analysis is steadily becoming more central to management decision-making in media organisations across the globe. The reliance of SVoD services such as Netflix on data analytics to underpin decisions about new content investment is well established. However, what are the key opportunities and challenges facing the rest of the television industry? This article examines how data analysis is facilitating improved methods of personalisation and more effective intelligence about the relative appeal of content for differing audience segments. But growing reliance on big data also raises a number of critical public interest questions. This article highlights how data is now a key source of competitive advantage in the television industry and a resource that can be monopolized. It argues that media policy-making needs to pay more attention to the emergence and implications of asymmetries of power in relation to ownership and use of data in managerial decision-making.

Keywords:

Big data; data analysis; algorithms; genomes; personalization; monopolization.

Author contact details: Professor Gillian Doyle, Centre for Cultural Policy Research, University of Glasgow, 13 Professor Square, Glasgow, G12 8QQ. Gillian.Doyle@Glasgow.ac.uk. +00 44 141 330 3806.

Television and the Development of the Data Economy: Data analysis, Power and the Public Interest

Introduction

Data analysis is steadily becoming more central to management decision-making across the media and in the television industry (Athique, 2017; Couldry and Turow, 2014). Ongoing increases in the data yielded by online consumption of content provide numerous opportunities for television companies and promise to bring ever-closer alignment between the taste profiles of individuals across the globe and a growing universe of television content properties (Cohn, 2016). But capitalizing on such opportunities while safeguarding against threats to competition and to audience welfare calls for a better understanding of the ways that access to data and techniques of data analysis can underpin contemporary decision-making in the television industry.

This article examines how, for major broadcasters and other television content suppliers, analysis of digital data is playing an increasingly central role in providing insights about which properties will appeal to what audience segments and in which territories. The reliance of online SVoD services such as Netflix on data analytics to underpin decisions about new content investment is well established. However, what are the key opportunities and challenges facing the rest of the television industry as digital data analysis moves from the fringes to the forefront in processes of decision-making? Drawing largely on secondary sources (emerging literature, industry reports, and academic and practitioner presentations at recent television industry and policy workshopsⁱ), this article asks how the use of data analysis is being developed by television companies and it assesses the critical issues raised for public policy-makers, focusing particularly on emerging asymmetries of power and information.

From Fringes to Foreground

The television industry is beset by many changes at present and whereas, historically, power has tended to reside principally at the distribution end of the supply chain, now, because of the rise of web-connected globalized delivery formats and platforms and because of heightened competition in the distribution stage, the emphasis on vertical cross-ownership ownership of distribution *plus* content has increased, as evidenced by many recent takeovers (Esser, 2016; Doyle, 2017). But harnessing the power of big data has become a key driver of patterns of investment and of structural configurations in the television industry too. Indeed, some argue that, for media businesses, having access to data is now just as important as owning content and/or distribution (Arsenault, 2017).

While the challenges of transitioning from broadcasting to digital multi-platform content delivery are manifold, one major beneficial by-product of digital distribution for television companies has been the introduction of return path data. Every viewing event on web connected platforms creates data that can be analysed (Pennington, 2017). And as distribution of media content has migrated to web-connected platforms and devices, media suppliers have become increasingly focused on harnessing the value of digital traces left by consumption of that content (Napoli, 2014).

Recognition of the impact that big data is having on businesses and society has impressed itself on recent academic literature. Studies from information management and business studies have sought to illuminate how, across many sectors of the economy, big data and advances in data analytics are reshaping business models and employment (Loebbecke and Picot, 2015) and offering companies

opportunities to profit (Schweidel, 2015). In studies of management of media, the ‘algorithmic turn’ has stimulated work on how companies can use digital data to better manage their interactions with audiences and to derive new revenue streams. For example, Evens and Van Damme (2016) have highlighted how news organisations can profit from audiences’ willingness to relinquish personal data in return for access to content, thus enabling a diversification away from traditional sources of advertising income which have been in decline.

A number of earlier academic studies have traced the rise of ‘algorithmic culture’ in the television industry (Cohn, 2016). Much work has focused on how big data has transformed audience research (Athique, 2017). Some has examined, for example, how the information that emerges from measuring television audience engagement on digital platforms differs from traditional television ratings (Kosterich, 2016). Many studies have examined the development and implications of systems of recommendation and, in particular, Netflix’s algorithm-based approach (Hallinan and Striphias, 2014). However, as Napoli argues, relatively little research has been produced so far ‘exploring how big data are being utilized in the management of media organisations’ (2016: 3). This article extends knowledge by examining current areas of investment in data analysis by UK broadcasters.

The potential for big data to generate insights about what audiences like is of central interest to television companies. It is well known that production of cultural outputs and especially television content involves high costs and risks while, at the same time, ‘nobody knows’ what will succeed with audiences (Caves, 2000). This state of affairs has encouraged a long history of attempts to quantify likely success factors for cultural outputs. For example, earlier research has attempted to quantify the impact of different cinematic components on a film’s subsequent critical and/or box office success (Simonton, 2011; Ginsburgh and Weyers, 2014). The arrival of the internet has greatly helped such research by placing in public domain large amounts of data about the characteristics and performance of cultural outputs. For example, the Internet Movie Database (IMDb) is an online database that provides information about films, television programmes and video games, including about their cast and production crews plus user reviews. Recent exploratory research at the University of Warwick suggests that the success of a television programme can be inferred with reasonably high levels of accuracy from analysis of data readily available on IMDb about a small number of key traits, namely: genre, country of origin and patterns of release (Parry, Pogrebna and Del Vecchio, 2017).

Thus data analysis, traditionally regarded with some wariness in the television industry and confined to Information Technology (IT) departments, has stepped out of its silo and into the mainstream of management decision-making. However, this development poses new challenges for media companies and for consumers and public policy-makers.

One such challenge is how the endemic organizational culture of some media companies remains resistant to the integration of data specialists, even though the need to bring IT and digital knowhow into strategic decision-making may be well recognized (Doyle, 2015). Another strong challenge is that of calculating how to use data judiciously in everyday management decision-making and how to avoid ‘editing by numbers’ (ibid: 57). In the television industry, the question of how far algorithms should be allowed to guide content recommendation systems has received considerable attention (Calvado, 2017). While the speed and accuracy of computers in processing large datasets makes reliance on algorithms tempting, the evidence of earlier research highlights the dangers of sidelining human judgment entirely in favour of computational methods (Lewis, Zamith and Hermida, 2013: 49). And a further challenge for media organisations is how to evolve systems of data management that comply with relevant legal frameworks which, in some cases at least – e.g. in the EU which is transitioning to the General Data

Protection Regulation in May 2018 (Bourreau, de Streel and Graef, 2017) - are undergoing change in response to public concerns about the potentially harmful implications of large scale data-mining.

The use of data by media companies is, of course, nothing new but digitization has fuelled a significant rise in reliance on data analysis to personalize media content offerings and to inform content investment decisions. Although Cohn (2016: 678) cautions that ‘the role that users have in reading and decoding the culture around them’ should not be under-estimated, the extent to which consumers are fully aware of how their everyday activities online leave behind digital traces and how these are used by digital content service providers, marketers or others is debatable. While data analysis and use of algorithms *is* associated with a number of benefits for users, especially lower transaction costs, fewer information problems and speedier decisions, it also raises a number of welfare concerns related to privacy, surveillance and pluralism (Rubinfeld and Gal, 2016; Kennedy and Moss, 2015).

Albeit that, as Athique points out, that ‘those outraged by the privacy implications of what is essentially a world-spanning phone tap operation are commonly reminded of the old adage: “If its free then you’re the product” (Athique, 2017: 5), the hidden costs of supposedly free goods may not always be apparent to consumers. From the perspective of public policy-making, weighing up the benefits of effective targeting and lower prices for online services against the adverse effects of loss of privacy and/ or the nuisance of intrusive advertisements involve trade-offs that are complex (Bourreau, de Streel and Graef, 2017).

Some critics also point to the effects that personalization may have on shared public discourse. Couldry and Turow (2014) argue that the advertising industry’s use of big data to target advertising, through fuelling more personalized content, is fracturing audiences and eroding shared civic culture. A growing reliance on big data analysis is, they argue, ‘clearing’ the landscape of one of the ‘basic ingredients of democratic life: the reliable and regular exchange of common ideas, facts, and reference points about matters of common concern’ (Couldry and Turow, 2014: 1722)

While the increasingly prevalent use of data does clearly raises a number of critical public interest questions, particularly surrounding privacy, this article argues that a further pressing issue which deserves more recognition than it has received to date is the emergence and implications of asymmetries of power in relation to ownership and use of big data in media decision-making.

Uses of Data Analysis

‘Because we have a direct relationship with consumers, we know what people like to watch and that helps us understand how big the interest is going to be for a given show...like *House of Cards*’

(Friedland, Chief Comms Officer of Netflix cited in Carr, 2013.)

Whereas SVoD services such as Netflix have led the way in using data analysis to assess the likely audience reception for new television shows, now the television industry as a whole is following suit by seeking to harness data in order to inform content decisions (Observation, Salford: 26 January, 2017). This is evidenced by the way that, for example in the UK, broadcasters have moved towards making registration compulsory for viewers in order to access online catch-up services (Clark, 2017). Registration means providing at least some basic personal information that enriches the data about individual viewing experiences that the broadcaster can collect and use. In the commercial sector, mainstream free-to-air services such as ITV and Channel Four already require registration before

giving access to their on-demand services *ITV Hub* and *All 4 On Demand* respectively. In announcing that it plans to make registration compulsory for users of the *iPlayer* from 2017, the BBC made clear that this would benefit audiences by facilitating more personalization. According to Director General Tony Hall:

‘I want everyone to get the very best from the BBC. By learning about what you want and like we can take you to more of the great programmes you love, stories you might be interested in and content you might otherwise never have discovered’

(Hall, cited in Ling, 2016.)

Across the board, it is evident that not only maturing SVoD platforms but also traditional television broadcasters with their well-stocked catch-up services are focused on how best to use data analysis to enable viewers to make better informed choices. Since abundance is an intrinsic feature of digital on-demand services, the appeal of any offering depends not only on having content that audiences like but also on providing navigation tools that make choices manageable (Calvado, 2017). For many on-demand content services, the use of recommendation systems - which typically are based on analyzing data about the sort of content that users with similar taste patterns have chosen in the past in order to predict material that a current user will like – provide a solution. As broadcasters evolve their on-demand content services, the development of data-informed systems of profiling that facilitate processes of navigation and selection through personalized recommendations and tailored homepages has become a priority (Observation, Salford: 26 January, 2017).

This approach underpins, for example, Spotify’s *Discover Weekly*, a tool that uses algorithms in order to provide service users with a new personalized playlist every week based on analysis of the consumption behaviours of others with similar tastes (Observation, Salford: 26 January, 2017). The development a personalized playlist, whereby ‘millions of people [are] treated as uniques as opposed to millions of people being treated as one of many’ (Page, cited in Simms, 2016), was a major innovation as a curation tool when it was launched in July 2015 and many Spotify subscribers are now habituated users of this facility according to Spotify’s Director of Economics Will Page (Observation, Salford: 26 January, 2017).

The challenges for on-demand services that offer audiovisual content as opposed to music are distinct because of immense difference in the relative sizes of television versus music catalogues and because of the differing ways that music listeners, as opposed to television viewers, select and consume content. But, even so, it is conceivable that, as data about individual viewing choices becomes more abundant over time, television content providers may eventually emulate Spotify in drawing on collaborative filtering to offer a bespoke individualized curatorial service to users of their content archives.

At present however, navigation and cross-promotion of television content are centred around recommendation systems and most if not all broadcasters are investing in improving the accuracy of these (Observation, Salford: 26 January, 2017). A key challenge television service providers face is the trade off between, broadly, algorithmically based filtering systems that are relatively low cost versus more complex and tailored ones that require more human curatorial input. According to Greg Detre, Chief Data Scientist at Channel Four, whereas automated recommendation systems will often come up with suggestions that audiences find suitable, they are generally weighted towards ‘raw popularity’ (Detre, 2017). However, broadcasters’ needs for a more refined means of targeting and developing audiences is propelling new investment into research and development of increasingly sophisticated filtering systems based not only on automated metrics but also on human editorial input (Detre, 2017; Observation, Salford: 26 January, 2017)

Several US and UK broadcasters have gone beyond investing purely in recommendation engines to, in addition, creating systems of analyzing data that help with predicting which forms of content will have appeal, and for whom (Observation, Salford: 26 January, 2017). For many, this involves developing ‘genomes’ whereby content is broken down or classified into very detailed categories (such as the production location or the mood of the script) so that how attractive each constituent component is for differing audience segments can then be subjected to close analysis (ibid). This may involve, as a starting point, testing viewer responses to content and applying predictive analytics and multiple regression models to tease out which of the component characteristics of television programmes tend to engender the strongest reactions from audiences (Barsotti, 2017).

The ability to harness data analysis in order to predict what sort of content will appeal to which audience segments and in which territories has increased enormously in the digital era thanks to not only the availability of massive amounts of audience data but also on account of the rise of automated systems for identifying and classifying content by traits (Observation, Salford: 26 January, 2017). A number of broadcasters are currently working on developing automated systems whereby machine readings of synopses and descriptions of television programmes provide a basis for assigning items of content with hundreds of contextual attributes, for example: genre and sub-genre; nationality of actors; location; mood of the plot; whether music plays an important role in driving the narrative, etc (Danaher, cited in Barsotti, 2017). Armed with a detailed ‘genome’ for each title in their portfolio, owners of television content portfolios can then tease out recurrent trends in relation to what sort of content has appeal and for whom.

This approach follows in the footsteps of earlier work in the music industry, such as Music Genome Project, which involves analyzing and classifying songs according to a framework of hundreds of attributes in order to identify traits that are popular and to find and recommend works that contain similar elements (Kumar, Sowmya and Srinivasa, 2016). At the forefront of such approaches in the context of the television industry is BBC Worldwide, the commercial arm of the BBC, where the Insight Team, led by David Boyle, is developing the use of genomes and predictive analysis in order to help forecast which BBCW content properties are likely to appeal in differing international markets (Boyle, 2016). According to Boyle, ‘data helps the BBC determine which countries a new show is best suited for—and which short-form content will be most engaging in promoting those shows’ (ibid). The BBC is far from alone in seeking to develop systems for classifying and coding content according to a myriad of traits in order to then predict their likely appeal for differing audience segments – commercial players such as Sky in the UK and HBO and Disney in the USA have also been investing in developing systems that more effectively optimize the match between their content, audiences and advertisers. Strategic takeovers offer an effective means of procuring specialist capability rapidly, as evidenced by, for example, the acquisition by Hulu (owned by NBC Universal, Disney and Fox) of Video Genome Project (VGP) in December 2016 (Goikhman, 2017; Spangler, 2016) or by Sky’s investment in technology companies such as DataXu (McDonald, 2016).

Thus, in the television industry, the race is on to develop the most effective systems of anatomizing and classifying content in such a way as maximizes its predictive value. Harnessing big data in this way provides value to television companies by (a) yielding insights about the relative appeal of content which, in turn, may help to inform pricing and sales negotiations surrounding finished product; and (b) offering intelligence as to what sorts of content are worth investing in the first place – a benefit that SVoD services such as Netflix already exploit to evident advantage.

The Development of the Data Economy and The Public Interest

While the ways that big data can inform and facilitate decision-making undoubtedly promise benefits for the television industry and for audiences, the migration towards greater reliance on data analysis does also raise a number of public interest concerns. In particular, a key concern for policy-making is the emergence and implications of asymmetries of power in relation to ownership and use of data which now represents a key informational resource in the television industry. As is argued below, the growing centrality of data analysis raises other concerns too related to widespread use of personalization techniques and to the future of content investment.

Informational Asymmetries and Market Power

Whereas in many other sectors of the economy (e.g. insurance) an extensive reliance on statistical analysis to drive business decisions is nothing new, for media, digital distribution has been a catalyst for the inexorable rise of data analysis (Observation, Salford: 26 January, 2017). Television executives might have bragged in the past that investment in hit shows was down to ‘gut instinct’ but, in the digital era, wisdom distilled from big data is unquestionably the aid of choice. Data analysis is also indispensable in strategies of targeting and addressing digital audiences. For media companies, Google’s Chief Economist Hal Varian’s observation (cited in Lohr, 2009) that statistical analysis would be ‘the sexy job in the next ten years’ has turned out to be highly prescient.

Harnessing data effectively requires investment in such activities as recording, collecting, storing and analyzing data to extract value (Mayer-Schonberger and Cukier, 2013). In the television industry, broadcasters and other content service providers have made substantial investments in these activities based on the understanding that the recording or creation of datasets simultaneously brings *ownership* of that data and the right to exploit it. Access, on an exclusive basis, to the valuable information that can be drawn out of proprietorial datasets, enables a return on investment to be earned. As Brobst, CEO of data analytics firm Teradata puts it:

‘Let’s say I’m Netflix and I collect lots of data ... to provide a service to you, it shouldn’t be that my competitors should benefit from all this investment I’ve made... That’s my asset’
(Brobst, cited in Wolpe, 2015)

In effect, ownership of or access to data and to the techniques needed to make sense of data so as to inform management decision-making are now key sources of competitive advantage in the television industry and they are resources that can be *monopolized*.

In drawing attention to the need for competition authorities to take heed of this and head off exclusionary and anti-competitive practices, some have argued that data should be seen as ‘the new oil’ in the economy (Economist, 2017). This analogy is misleading in that, unlike oil, which is a scarce physical commodity, data is intangible and it is non-rivalrous in that its supply is not diminished through usage (Schepp and Wambach, 2016). But data, like oil, constitutes an important input factor to the economy in the digital era. At the same time, data-gathering and analysis are characterized by the widespread availability of economies of scale and scope which means that the economics of these activities ‘favour market concentration and dominance’ (OECD, 2014: 7).

In media industries, where the strategic importance of access to relevant data is especially great, the case for regulatory oversight to avoid potentially anti-competitive effects is compelling. As Arsenaault suggests, the need for media companies to control specific distribution mechanisms has waned nowadays next to the importance of ‘leveraging data to predict which delivery mechanisms work best

for which product, at which particular moment' (Arsenault, 2017: 20).

Earlier research has, rightly, drawn critical attention to the potent threats to pluralism, democracy and social cohesion posed by unrestrained monopolization in media industries (Freedman, 2012; Iosifidis, 2011). But monopolization also affects economic efficiency. Monopolisation of data militates against the most efficient utilization of what is an 'infrastructural resource' (OECD, 2014: 5) which can, theoretically, be re-used by many different companies bringing increasing returns to scale and scope, growth in productivity and resultant broad economic and societal gains. But however strong the efficiency-based arguments against exclusion might be, it is unsurprising that monopolisation occurs, given the substantial investments required 'to develop and maintain databases, meta-data and related algorithms' (ibid: 6).

While uneven knowledge amongst competing suppliers about what audiences like is a long-standing fact of life in media markets, the informational asymmetries that have emerged alongside the rise of globalized online distribution have raised concern, not least on account of the growing dominance of a handful of US-based technology driven players such as Netflix and Amazon, in amassing and drawing value from personal data (McLaughlin, 2017). Concerns about monopolization have prompted policy-makers in the EU to question the effects of such dominance both in relation to consumer privacy and to competition and innovation (Van Gorp, 2015; Bourreau, de Streeel and Graef, 2017; De Streeel and Herrera-Gonzalez, 2017).

Netflix, which has become an increasingly important commissioning purchaser of original television content, has been criticised for hoarding data and 'refus[ing] to reveal figures on how popular, or not, its shows are' (Sweney, 2017). One special concern is the resultant informational asymmetries which may skew negotiations between television service providers (buyers) and producers (sellers) of content to the detriment of the latter. It is well recognized that in order to build their businesses, television producers need to negotiate favourable terms on retention and secondary exploitation of the rights to their content assets. Producers' ability to do this is dependent on the distribution of bargaining power along the vertical supply chain (Doyle, 2017: 7) but also, crucially, it rests on producers having information about the popularity of their own shows. Whereas traditionally producers, broadcasters and other distributors have been able to draw on readily available and identical audience data from third party sources such as BARB or Nielsen, producers whose output is commissioned by SVoD players are now frequently faced with the problem of sparse or non-existent access to information about how their content has fared with audiences.

It is well recognized that informational asymmetries may, at times, result in market failure. Economics Nobel Laureate Akerlof famously used the example of transactions in used cars or 'lemons' to demonstrate how when buyers or sellers do not have the same information this can lead to market failure and to an inefficient use of resources (Akerlof, 1970). Likewise if distributors and producers of television content lack correct information about the popularity of specific content propositions then, theoretically, it is more likely that investment in provision of said content will take place at a level that is greater or less than is socially optimal.

But whether, in reality, monopolization of data and related problems of asymmetric information might provide a market-failure based justification for intervention is debatable (Cowen and Crampton, 2002). The relationship between monopoly and technological innovation has never been straightforward. While some economists argue that monopolists hold back rivalry and therefore deter innovation, others adhere to the more Schumpeterian viewpoint that the incentive of being able to reap monopoly profits, at least in the short term, is precisely what encourages firms to make risky investments in new

initiatives in the first place (Doyle, 2013: 185). Few would deny that those who invest in collecting and analyzing data in innovative ways so as to better serve audiences and build their businesses deserve rewards. But that the possibilities for analytics-driven innovation which have excited imaginations right across the television industry can only be realized by those handful of broadcasters and SVoD companies who enjoy access to the *de facto* means of participation in harnessing digital datasets seems, from a wider societal perspective, somewhat wasteful. Of particular note is the fact that ‘independent’ television production companies who are *not* vertically integrated into broadcasting or distribution entities with relevant resources will be left behind in developing and exploiting data-insight driven approaches to investment in new content. This informational asymmetry exists at the very time when UK and European policy-makers are extolling the importance for our creative production sectors of using new technology in order to prosper (Create UK, 2014).

Personalisation

Another area of concern surrounding the developing use of data analysis by television companies relates to the widespread collection and use of personal data which has drawn criticisms about privacy and about the effects of an ever-growing and pervasive emphasis of personalization. Despite concerns about the implications for privacy of large scale data mining, it is broadly well-recognised that a trade off exists between privacy and lower prices to the consumer for online services – essentially data disclosures help to pay for content services where otherwise charges would have to be levied on the user (Deutscher, 2017; Evens and Van Demme, 2016). Likewise with personalization techniques based on algorithmic analysis of personal datasets, while its attendant acquiescence to the creep of robotics (Gal, 2017) and its normalization of large scale data mining (Cohn, 2016) may well be concerns, at the core lies a trade-off between data disclosure and the benefit of recommendation systems that audiences generally tend to regard as helpful and convenient.

Clearly personalization helps television audiences to navigate vast archives of content. However, as marketing analysts have discovered, ‘[g]reater personalization typically increases service relevance and customer adoption, but paradoxically, it also may increase customers’ sense of vulnerability and [therefore] lower adoption rates’ (Aguirre et al, 2015). So for television companies a key challenge is the decision about ‘where to draw the line between the usefully personalized and the excessively personal’ (Guszeza in Schweidel, 2015: xvii). In announcing that registration to use the iPlayer will become compulsory in 2017 as part of a ‘reinvention’ of PSB for the digital age, Director General Tony Hall argued that ‘[m]illions of people are already benefitting from this more personalized BBC and by rolling it out for everyone no one will be left behind’ (Hall, cited in Ling, 2016). But the BBC has allowed that those registering may forego personalisation by adjusting their privacy settings accordingly. Such sensitivity to public ambivalence about data mining and increased levels of surveillance on the part of a publicly-funded broadcaster is clearly prudent but not all television services make it so easy to opt out of personalization techniques.

Another peril of recommendation systems is their capacity to ensnare viewers in filter loops and bubbles at the expense of potential for serendipity and taste development. Earlier work has highlighted the dangers of filter bubbles in relation to online news content where, confronted by a multitude of offerings, audiences may end up ‘dwell[ing] intellectually only in arenas where they are comfortable, creating barriers between them and those in the public sphere with whom they would not likely agree’ (Couldry and Turow, 2014: 1711). Filter bubbles, as they affect news consumption, are especially worrying because access to an array of political viewpoints is seen as part of the bedrock of democracy. But access to a plurality of entertainment content also matters to individuals and to societal aspirations relating to promotion of tolerance and socio-cultural cohesion. So, as television

companies invest in the development of algorithmically generated recommendation systems, the knowledge and skills needed to interrogate the design of such systems and to consider, for example, the extent to which algorithms are constructed to achieve easy hits or by contrast to slightly challenge and develop existing taste profiles, will become increasingly relevant to the domain of media policy-making.

Implications for Local Content

A key focus for investment at present is development of genomes or systems for anatomizing television content in order to better predict its likely appeal in differing markets. As discussed earlier, the application of techniques based on machine reading and categorization of content to create detailed genomes is very much aimed towards understanding the wants and needs of *international* audiences. To the extent that insights gleaned about content that has wide international potential may, over time, increasingly influence content investment decision-making, a correspondingly negative implication for local content is inescapable.

Does it matter if data-informed content investment decisions favour material that is internationally-oriented over content that serves the narrower interests of local and national television audiences? Television is widely recognized as a key aspect of the landscape of everyday cultural provision and as a vector through which societal ‘values, myths and symbols that serve to unite’ are shaped and shared (Smith, 1998: 187). As such, the availability of content aimed primarily at local audiences and the potential threat to diversity and societal welfare posed by marginalisation of local content are concerns that surface regularly in media and cultural policy-making discourses. Referring to the disruption caused by new digital intermediaries such as Netflix the Canadian Heritage Minister, Mélanie Joly, recently warned that that ‘[i]f we’re complacent, this new wave of information can drown out our own content—our French-language TV and films; our indigenous music; our multicultural programming’ (Joly, 2017).

It is widely recognized that sustaining diversity has become much more difficult with ‘the emergence of powerful transnational platforms commercialising cultural goods and services online’ (García Leiva and Albornoz, 2017:10). Even so, as is reflected in a plethora of national and international policy initiatives such as the UNESCO Convention on Cultural Diversity, preservation of access to diverse audiovisual outputs, including indigenously-made television content, is regarded as essential in promoting social cohesion and integration, in maintaining plurality and democracy and in sustaining regional and national identities and in some cases languages (UNESCO, 2005).

The development of profiling techniques to help identify material likely to appeal in international markets has implications not only for content but also for content production. Earlier research has drawn attention to the critical industrial issues surrounding increased reliance on international markets as the primary source of financing for production of screen content (Paterson, 2017). Any push towards content that works predominantly for international markets makes it more difficult for programmes geared mainly towards the concerns of domestic audiences to attract production funding (Esser, 2016: 3605). So, at stake here is not only the welfare of local audiences but also, potentially, the economic position of local producers in an increasingly globalised television environment.

Conclusions

The rise of digital distribution has provided television companies with unprecedented levels of data and, in turn, enabled development of improved tools and methods for understanding the tastes, preferences and behaviours of their audiences. For UK broadcasters, a major focus for investment is on improved data-informed systems of profiling to facilitate easier navigation and more effective curation of content (Observation, Salford: 26 January, 2017; Pennington, 2017). A move by some traditional broadcasters, such as Channel Four, towards blending manual with computational methods reflects the fact that, whereas full automation may be attractive on grounds of cost, the integration of human input that brings contextual sensitivity and rigour alongside machine based computation is apt to achieve optimal results (Lewis, Zamith and Hermida, 2013).

Not only can analysis of online consumption data provide a closer understanding of the taste profiles of audiences at individual and aggregate level, but also, as systems for tagging and coding content become increasingly sophisticated, data analysis can yield ever more detailed insights into the sort of component characteristics of content that are likely to appeal to audiences (Observation, Salford: 26 January, 2017; Parry, Pogrebna and Del Vecchio, 2017). Another key area for investment by television companies at present, especially for companies such as the BBC that are major owners of content, is the development of genomes or automated systems of classifying content according to its constituent traits with a view to then teasing out what sorts of content will appeal to which audience segments and in what territories. Thus, lured by the promise of a cornucopia of market intelligence, data analysis has moved from the fringes to the foreground as an area for investment in the television industry.

The migration towards ever-increasing reliance on data analysis has triggered a number of policy concerns and this article calls for a refocusing of regulatory priorities around use of data. While privacy has been the main focus for intervention to date, this article argues that media policy-making needs to pay more attention to the emergence and implications of asymmetries of power in relation to ownership and use of data which now represents a key informational resource in the television industry.

Earlier research has drawn attention to the ‘growing significance of data as an input factor and valuable asset’ and the need for policy-makers to try and ensure that concentrated control over data does not result in market distortions (Schepp and Wambach, 2016: 124). This is especially true in television industry where opportunities to profit, both commercially and creatively, from what Hallinan and Striphas (2016: 117) describe as ‘render[ing] algorithmic information processing systems legible as forms of cultural decision making’, abound. It is well-established that decisions about how data is used can have significant welfare implications (Thornham and Cruz, 2017). So, as well as striving to promote fair competition and efficiency, data-related policy interventions need to reflect additional societal aspirations impinged upon by the use of data in managerial decision-making in the television industry such as diversity and taste development.

However, given the increasing complexity of the globalised digital media distribution ecosystem, policy-making in this area is, as pointed out by the OECD (2014: 7), bedeviled by considerable challenges in relation to defining relevant markets, pinpointing levels of market concentration and quantifying exact welfare implications. Addressing such challenges requires appropriate expertise. Therefore if the public interest is to be safeguarded effectively at the same time as industry strives to harness the advantages of changing technology, it is not only television companies but also media and competition regulators who now need to increase their investment in specialist skills in data analysis suited to the challenges of big data era.

Acknowledgements:

The generosity of the organisers of two key events at which participant observation was carried out - the *Data Science in the Entertainment Industries* workshop held in Manchester on January 26-27 2017 and the *Economics, Law and Policy of Communications and Media* seminar held at EUI, Florence, Italy on March 24-5 2017 – is gratefully acknowledged. The support of the UK Economic and Social Research Council (ES/N015258/1) is also gratefully acknowledged.

References:

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K and Wetzels, M (2015), Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness, *Journal of Retailing*, 91 (1): 34–49.

Akerlof, G (1970), The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism, *The Quarterly Journal of Economics*, 84 (3): 488-500.

Arsenault, A (2017), The datafication of media: Big data and the media industries, *International Journal of Media and Cultural Politics*, 13(1-2): 7-24.

Athique, A (2017), The dynamics and potentials of big data for audience research, *Media, Culture & Society* (early online publication)

Barsotti, S (2017), ‘Hold tight and pretend it’s a plan: Big Data and BBC’s Doctor Who’, *Carnegie Mellon University News*, February 21, 2017. Accessed at: <http://www.heinz.cmu.edu/news/news-detail/index.aspx?nid=3925>

Boyle, D (2016), Catchy content: What makes TV content work?, *Strata Data Conference*, New York, Accessed at <https://conferences.oreilly.com/strata/strata-ny-2016/public/schedule/detail/53526>

Bourreau, M., de Streel, A., and Graef, I (2017), *Big Data and Competition Policy: Market Power, Personalised pricing and Advertising*, Brussels: Centre on Regulation in Europe (CERRE).

Calvado, E (2017), ‘Recommender Systems: Trust and Biased Advice’, *Economics, Law and Policy of Communications and Media*, FSR Communications & Media, CMPF and ENTrANCE Annual Scientific Seminar, Florence: EUI, March 24-5 2017.

Carr, D (2013), Giving Viewers what they want, *New York Times*, February 24, 2013. Accessed at: <http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?mcubz=0>

Caves, R (2000), *Creative industries: contracts between art and commerce*, Cambridge, MA: Harvard University Press.

Clark, L (2017), You’ll soon need an account to watch BBC iPlayer, *Wired*, 11 May 2017. Retrieved at: <http://www.wired.co.uk/article/bbc-iplayer-to-require-login-from-2017>.

Cohn, J (2016), My TiVo Thinks I'm Gay: Algorithmic Culture and Its Discontents, *Television and New Media*, 17(8): 675-690.

Couldry, N and Turow, J (2014), Advertising, Big Data and the Clearabnce of the Public Realm: Marketer's New Approaches to the Content Subsidy, *International Journal of Communication*, 1710-1726,

Cowen, T and Crampton, E (2002), *Market Failure or Success: The New Debate*, Cheltenham, UK: Edward Elgar Publishing.

Create UK (2014), *Creative Industries Strategy*, London: Creative Industries Council UK. Accessed at: www.thecreativeindustries.co.uk/media/243587/cic_report_final-hi-res-.pdf

Detre, G (2017), Centaur Evaluation, Combining Minds and Machines for Better Recommendations, Presentation at *Data Science Conference*, BBC Salford, January 26-27, 2017.

Deutscher, E (2017), 'The Role of Personal Data and Privacy in Merger Analysis', *Economics, Law and Policy of Communications and Media*, FSR Communications & Media, CMPF and ENTraNCE Annual Scientific Seminar, Florence: EUI, March 24-5 2017.

Doyle, G (2015), *Understanding Media Economics* (2nd edition), London: Sage Publications.

Doyle, G (2015), Multi-platform media and the miracle of the loaves and fishes, *Journal of Media Business Studies*, 12 (1): 49-65.

Doyle, G. (2017) Television production: configuring for sustainability in the digital era. *Media, Culture and Society*, Early Online Publication.

De Streel and Herrera-Gonzalez, F (2017), 'How to integrate Zero-priced Markets in Anti-Trust Analysis', *Economics, Law and Policy of Communications and Media*, FSR Communications & Media, CMPF and ENTraNCE Annual Scientific Seminar, Florence: EUI, March 24-5 2017.

Deutscher, E (2017), 'The role of personal data and privacy in merger analysis – a critical assessment of the EU Commission's Facebook/ WhatsApp decision, Competition, Regulations and Freedom of Expression', *Economics, Law and Policy of Communications and Media*, FSR Communications & Media, CMPF and ENTraNCE Annual Scientific Seminar, Florence: EUI, March 24-5 2017.

Economist (2017), The worlds most valuable resource is no longer oil, but data, Leader article in *The Economist*, May 6, 2017, Accessed at: <http://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>

Esser, A (2016), Challenging US leadership in Entertainment Industries? The Rise and Sale of Europe's International TV Production groups, *International Journal of Communications*, 3585-3614.

Evens, T and Van Damme, K (2016), Consumers' Willingness to Share Personal Data: Implications for Newspapers' Business Models, *International Journal on Media Management*, 18 (1): 25-41.

Freedman, D (2012), Metrics, Models and the Meaning of Media Ownership. *International Journal of Cultural Policy*, 20 (2): 1-16.

Gal, M (2017), 'Algorithmic Consumers', *Economics, Law and Policy of Communications and Media*, FSR Communications & Media, CMPF and ENTraNCE Annual Scientific Seminar, Florence: EUI, March 24-5 2017.

García Leiva, M and Albornoz, A (2017), Cultural industries and diversity: old debates and new challenges, *Quaderns del CAC*, 43, vol. XX: 5-15.

Ginsburgh, V and Weyers, S (2014), Nominees Winners and Losers, *Journal of Cultural Economics*, 38(4): 291-313.

Goikhman, D (2017), How big data will shape the hypercompetitive online video space, January 20, 2017. *CMSwire.com*, Accessed at: <http://www.cmswire.com/digital-experience/how-big-data-will-shape-the-hypercompetitive-online-video-space/>

Hallinan, B and Striphas, T (2016), Recommended for you: the Netflix Prize and the production of algorithmic culture, *New Media & Society*, 18(1): 117–137.

Iosifidis, P (2011), *Global Media and Communication Policy: An International Perspective*, Basingstoke: Palgrave MacMillan.

Joly, M (2017), *Launch of Creative Canada*, Ottawa: Speech to Economic Club of Canada by the Minister of Canadian Culture, September 28. Accessed at: https://www.canada.ca/en/canadian-heritage/news/2017/09/creative_canada_-_a_vision_for_canada_s_creative_industries.html

Kennedy, H and Moss, G (2015), Known or Knowing publics? Social media data mining and the question of public agency, *Big Data & Society*, 2(2): 1-22.

Kosterich, A (2016) Reconfiguring the "Hits": The New Portrait of Television Program Success in an Era of Big Data, *International Journal on Media Management*, 18(1): 43-58.

Kumar, D, Sowmya, B and Srinivasa, K (2016), 'A comparative study of classifiers for music genre classification based on feature extractors', in *Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), IEEE* (pp. 190-194). IEEE.

Lewis, S., Zamith, R and Hermida, A (2013), Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods, *Journal of Broadcasting & Electronic Media*, 57:1, 34-52.

Loebbecke, C and Picot, A (2015), Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda, *The Journal of Strategic Information Systems*, 24(3): 149-57.

Lohr, S (2009), For Today's Graduate, Just One Word: Statistics, *New York Times*, August 6. Accessed at: http://unionstats.gsu.edu/8220/NYT_Lohr_StatisticsJobs.pdf

- Ling, T (2016), You'll need a password to watch BBC iPlayer from 2017, *Radio Times*, September 27 2016. Accessed at: <http://www.radiotimes.com/news/2016-09-27/youll-need-a-password-to-watch-bbc-iplayer-from-2017>
- Mayer-Schonberger V and Cukier K (2013), *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York: John Murray Publishers
- McDonald, A (2016), Sky buys into big data business, TBI, January 25, 2016. Accessed at: <http://tbivision.com/news/2016/01/sky-buys-big-data-business/537682/>
- McLaughlin, D (2017), Are Facebook and Google the new Monopolies? *Bloomberg*, 13 July 2017. Accessed at: <https://www.bloomberg.com/news/articles/2017-07-13/antitrust-built-for-rockefeller-baffled-by-bezos-quicktake-q-a>
- Napoli, P (2014), On automation in media industries: Integrating algorithmic media production into media industries scholarship, *Media Industries*, 1(1): 33–38.
- Napoli, P (2016), Special Issue Introduction: Big Data and Media Management, *International Journal on Media Management*, 18 (1): 1-7.
- OECD (2014), *Data-driven Innovation for Growth and Well-being: Interim Synthesis Report*, Paris: OECD, October 2014.
- Parry, G., Pogrebna, G and Del Vecchio, M (2017), 'The Science of Nudging Viewers Using IMDb TV Show Profiling', Presentation at *Data Science Conference*, BBC Salford, January 26-27, 2017.
- Paterson, R (2017), 'The competition discourse in British broadcasting policy', *CREATE Working Paper 2017/02*, Glasgow: CREATE.
- Pennington, A (2017), This time it's personal, *Broadcast TECH*, London: MBI, February/March 2017, pp16-19.
- Rubinfield, D and Gal, M (2016), The Hidden Costs of Free Goods: Implications for Antitrust Enforcement, *Antitrust Law Journal*, (80)3: 521-562.
- Schepp, N and Wambach, A (2016), On Big Data and its relevance for Market Power Assessment, *Journal of European Competition Law and Practice*, 7 2): 120-4.
- Schweidel, D (2015), *Profiting from the Data Economy: Understanding the Roles of Consumers, Innovators and Regulators in a Data-Driven World*, Upper Saddle River, NJ: Pearson Education.
- Simonton, D (2011), *Great Flicks: Scientific Studies of Cinematic Creativity and Aesthetics*, Oxford, OUP.
- Simms, D (2016), The state of music discovery: why Spotify's Discovery Weekly is the Music Industry's greatest Innovation of 2015, idobi, 26 April 2016. Accessed at: <http://idobi.com/editorial/the-state-of-music-discovery/>

Spangler, T (2016), Hulu acquires Video Genome Project to enhance recommendation for Live TV, VOD, *Variety*, November 15. Accessed at:

<http://variety.com/2016/digital/news/hulu-acquires-video-genome-project-vgp-1201918435/>

Smith, A (1998) Smith, A, *Nationalism and Modernism: A Critical Survey of Recent Theories of Nations and Nationalism*, London: Routledge.

Sweney, M (2017), Netflix tops 100m subscribers as it draws worldwide audience, *The Guardian*, July 18, accessed at: <https://www.theguardian.com/media/2017/jul/18/netflix-tops-100m-subscribers-international-customers-sign-up>

Thornham, H and Gómez Cruz, E (2018), Not just a number? NEETS, data and datalogical systems, *Information, Communication & Society*, 21(2): 206-321.

UNESCO (2005), *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*, Paris: United Nations Educational, Scientific and Cultural Organization, 20 Oct 2005.

Van Gorp, N (2015), *Competition among Information (Digital) Platforms, Study for ITRE Committee*, IP/A/ITRE/2014-11, Proceedings of the Workshop, 20 January 2015, Brussels: European Parliament.

Webster, J (2014), *The marketplace of attention: How audiences take shape in a digital age*, Cambridge, MA: The MIT Press.

ⁱ Participant observation was carried out at the *Data Science in the Entertainment Industries* Workshop led by David Boyle, EVP Insight at BBCW in Salford, Manchester on January 26-27 2017 and at the *Economics, Law and Policy of Communications and Media* Seminar, hosted by FSR Communications & Media, CMPF and ENTraNCE at EUI, Florence, Italy on March 24-5 2017.