



“Culture 4D: Digitization, Data, Disruptions, Diversity”

3rd Council of Europe
Platform Exchange on Culture and Digitisation

Tallinn, 29-30 September 2016

Digitization
Data
Disruptions
Diversity

From metadata to ‘big data’: Critical considerations on the emergent ‘digital memory industries

Indrek Ibrus PhD

www.culture4d.net



CHAIRMANSHIP OF ESTONIA
Council of Europe
May-November 2016
PRÉSIDENCE DE L'ESTONIE
Conseil de l'Europe



REPUBLIC OF ESTONIA
MINISTRY OF CULTURE



TALLINN UNIVERSITY

The opinions in this work are the responsibility of the author and do not necessarily reflect the official policy of the Council of Europe

Introduction

New large scale initiatives are emerging throughout Europe to digitise existing cultural resources, both tangible and intangible forms of cultural heritage. This will be a costly endeavour: for instance, the digitisation of European audiovisual heritage alone has been estimated at 5bn euros¹. The high costs are deemed worthwhile not only in terms of preserving the heritage, but also with regard to facilitating its new uses. In 2012 the EU Council suggested that the digitisation and online accessibility of Member States' cultural material were essential to: i) enable access for all to culture and knowledge in the digital era; and ii) ensure that digitised cultural material is available as an important resource for the European creative industries (Council of the European Union, 2012). Making this resource available for repurposing is expected to contribute to economic growth and job creation in Europe and the achievement of the EU's digital single market through the increasing availability of innovative online products and services. This has broadly been the plan. What I discuss in this paper are the related challenges, alternatives and implications for the public good in Europe.

Specifics of digital archives

Many researchers have distinguished between 'traditional archives' and network era archives². With traditional archives, everything collected and represented was selected by 'experts' according to some disciplined knowledge system. These experts thus managed the scope of society's dialogues with society's memory. This model was disrupted by the Internet, which facilitates self-organisation by consumers and their co-creation of archives and heritage content. Network era archives are all about experience of reception for audiences rather than an indication of official sanction³. However, the early era of networked archives is associated with insecurities for the involved parties (users, archivists, content authors). The productivity of online archives (such as YouTube or Flickr) is uncertain since their content is organised by agents with varying agendas and the archived objects may thus not retain their original integrity. As a result, responses to any queries by users tend to be unpredictable (Hartley, 2012). This leads to controversy regarding one of the particular benefits of audio-visual heritage digitisation: the opportunities for contemporary societies to have more immediate and therefore more 'open' relationships with their heterogeneous pasts (i.e. interpretations of visual materials can be more free and less dependent of dominant narratives, Torlasco, 2013). It is thought that access to digitised archives could give rise to a penumbra of 'new histories' that could affect broader reflective dialogues on existing genealogies and on 'digital historiography'. This has been the optimistic view (Ellis, 2012). In contrast, it is suggested (Baron, 2014) that amid the excess of accessible documents and their 'inappropriate' uses, historical awareness might diminish. Baron suggests that such a risk is effectively a 'natural effect' of digital archives.

1 Niggemann, de Decker, & Lévy, 2011

2 Ernst, 2013, 2015; Hartley, 2012

3 Baron, 2014; Uricchio, 2009

Cultural memory in the digital era

This relates to how the researchers of cultural memory in the era of networked culture talk about the evolution of their subject. Hoskins (2011), for instance, asks what are the prospects for the sharedness, stability and continuity of memory as it is increasingly connected with, newly ordered through and distributed across complex networks of digital media and technologies in our new memory ecology? He suggests that the digitally-enhanced paradoxes of flux and permanence, and immediacy and volume of access shape today's memory. Thus he talks about 'new memory', since in the era of networks memory is always 'new', given its continually emergent state availed through the metaphors, media and technologies of the day (ibid.). Ernst (2004) has suggested that instead of concerning ourselves with 'archival space' it is time to start talking about 'archival time' – referring to the 'dynamics of permanent data transfer' between the different archives, databases, media, modalities as well as cultures – the flux of contemporary archives is constant and, effectively, global. In this context Appadurai (2003) demonstrated the emergence of cultural memories for new virtual communities. He argues that the online archives do not present themselves as 'accidental repositories of default communities (like the nation)', but become deliberate sites for the production of anticipated memories by international communities.

Sketching a political economy of culture digitisation and metadata

While the described developments may appear to be historical forces too immense to be controlled, and while in aggregate they may be presented as complex in terms of effects, not just positive or negative, I would like to suggest a critical perspective scrutinising the techno-economic designs of the contemporary online archives, and their effects on how the cultural memories of contemporary media users may evolve. This is timely as there has been very little work within media and memory studies that takes a political economy approach to digital 'cultural memory industries'⁴. Yet, it is important, I suggest, to study the ways users can connect to their externalised memory resources to understand the effects of networks and digital infrastructures on ways memories are mediated and shaped.

Specifically, I would like to bring to the table the issue of metadata. When a cultural object – be it a film, TV programme, painting, photo, museum artifact, archival document etc. – is digitised then what makes it contextualisable, and therefore also searchable and findable, is 'metadata' – indexes and tags of various kinds – data about data. While it was suggested above that the global interconnectivity between archives and media is already happening, there are still major incompatibilities, especially between public and private archives and service providers. As an example, regarding metadata standards for audiovisual heritage, we are still in the typical early era of standards fragmentation: new *de facto* standards and methods are emerging around the world and this challenges cross-border co-operation and the technical interoperability of heritage databases. Yet, interoperable and freely available metadata is essential for the emergence of innovative services of which those associated with education are deemed to be the most immediately valuable. Further, it is the good quality of metadata and its seamless transferability across media that could alleviate the risk of content in user-led archives potentially losing contact with its original context and references, as discussed above.

4 Reading, 2014

Yet, reducing the incompatibility of archives and services tends to be undermined by the varying institutional needs for the functions of heritage metadata. I have shown that technical standards evolve through dialogue and power struggles among institutions with vested interests (see also Foray, 1995; Ibrus, 2013a, 2013b). Regarding audiovisual metadata⁵, while librarians have been concerned about standardised access to descriptors; producers are interested in efficient asset management (IPR, access controls); online service providers (YouTube, Netflix, etc.) are developing proprietary recommendations systems to secure customer loyalty and newly created dedicated public databases (e.g. Europeana, EUScreen) are seeking public value in service interoperability. The ‘multilevelled’ (Ibrus, 2013b, 2015) dialogues among these institutions are influencing the standardisation of metadata creation.

Yet, the challenge lies in those dialogues being often asymmetrical – some parties having significantly more bargaining power than others. While, for instance, the World Wide Web Consortium (W3C), together with European Broadcasting Union (EBU), Europeana and others, has been standardising new open standards to be used for videos in the ‘semantic web’ the dominant commercial service providers – Netflix, Amazon, YouTube and others – have chosen not to participate in these efforts. Instead, they are developing their own standards, producing their own metadata, which they do not share freely, but monetise in ways they see fit. Yet, these services matter as they dominate the markets. They are big because it is on their platforms that people or institutions, including often public archives, have chosen to share their content, either their own memories or remixes of others, be they originally private or public properties.

Effects of the ‘sharing economy’ on media concentration

In this context it is appropriate to bring in the ‘sharing economy’ concept. While this concept is often used to celebrate a more even distribution of agency in the economy, including media markets, it is also linked to the growing concentration in the global internet service markets.

What is the reason for the latter? Among the main conditioning factors is the phenomenon known as positive network externalities or network effects. The theory of network effects stems from economic theory and claims that the value of a network depends on the number of its users⁶. Hence, due to network effects, the more members a platform has for sharing purposes, the more attractive it is for its users⁷. Hence, large, international platforms for sharing purposes benefit from network effects that no national platform could offer. The problem with network effects is that this ‘value pull’ often leads to concentration in specific markets. In the domain of media and culture, concentration is feared due to its potentially negative effects on cultural diversity or political pluralism. Therefore, while a public archive or a rights owner for film heritage may be tempted to publish their holdings on Netflix or YouTube, due to their global reach they will not be able to control (or in case of Netflix even know about) the terms under which their content is found and accessed. There is evidence, for instance, that when memory institutions use YouTube to share their content, the latter’s algorithms suggest to users only the most popular videos and make finding the less popular, but still highly valuable, videos often impossible or at least the search results unpredictable⁸.

5 Wactlar & Christel, 2002

6 David & Greenstein, 1990; Katz & Shapiro, 1986

7 Cusumano, 2011

8 Vonderau, 2015

This means that such platforms, their algorithms and indexing systems have a crucial role in influencing what cultural resources can be found, and thereafter used for cultural memory construction. This suggests that these platforms, their technical designs, the standards they use and the rationales behind the choices of different standards and technologies all function ideologically, becoming part of the cultural construction apparatus.

Use of open metadata standards as a public policy objective

In this context I would like to point to the certain potential of article 3.5 of the Council of Europe's "Internet of Citizens" recommendation (2016). It argues for a new publicly available and sustainable digital space to be set up at European level, making use of existing European portals and platforms, to enable cultural resources and cultural knowledge to be legally shared and accessed without restriction of time and place. This digital space should provide, it is suggested, for a better global discoverability and accessibility of digital cultural resources and promote interaction, interoperability and collaboration among communities and between collections. It is also important, I would add, that this and other central platforms such as Europeana should continue to be as open and transparent as possible in terms of the standards, indexing techniques and technologies (including software and algorithms) they use – to encourage smaller European institutions to also choose open standards and to become aware of possible limitations to public value if they choose to co-operate extensively with proprietary and often less transparent service providers.

Emergence of cultural 'big data'

In addition to presenting cultural content on networked platforms more effectively – so that it is well contextualised and easily findable and reusable – there is also another concept of increasing importance that has implications for digital culture as a public good. This is data on how cultural content has been used on these platforms. Any click we make, any 'like' or comment we add, any recommendation we make to others – all is being recorded and used for various purposes – either for further personalisation of the service or for 'smart' marketing of various kinds of external products or services to users. This is where the concept of 'big data' comes in.

In the area of culture, 'big data' usually refers to born digital information that is user-generated and collected by computers. While the communicational model of online 'archives' such as YouTube or Instagram is often conceptualised as 'mass self-communication' (for distinguishing it from the forms of mass communication) their operational model is to invite continual input of data by individuals. As Couldry and Powell (2014) posit, "the exemplary product of mass self-communication is data". They explain that the economic model of mass media was structured around generating an audience whose attention could be sold to an advertiser. In the mass self-communication, on the other hand, model individuals are still part of an aggregate product, but instead of their attention on single messages, it is their own individual acts of communication that comprise the 'Big Data' and drive value-extraction. In relation to this, Puschmann and Burgess (2014) argue that the historical evolution of 'big data' is marked by a shift toward ever greater commercialisation of data.

Yet, what is often perceived as a challenge on this road is the actual manageability of big data – its harvesting, its processing, its sales under conditions of its quick growth, its varying quality and problems with its contextualisation and, therefore, also with its adequate analysis and uses. Puschmann and Burgess argue that the 'big data' trope often evokes the image of conjuring forces too powerful to control or curb. The implicit challenge, therefore, is how to control big data in order to successfully turn it into a resource.

But this works both ways – the usage of data as an extractable and tradable resource is similarly a challenge for civil society in terms of understanding whether this trade and the data uses run counter to public interest – harming not only the privacy of users, but perhaps more importantly the actual freedom of choice for all kinds of political, social and cultural agents. We know very little of how our data is being used, how it is commercialised and what the related effects are on the designs of all the information delivery services and life-co-ordination platforms that we consume. That is, as the phenomenon is in rapid development, and as such a ‘moving target’, there is a lack of rules, conventions or regulations aimed at securing transparent business conduct to make sure this conduct does not harm the principal freedoms of contemporary civil societies. On this front an important step was the recently adopted EU data protection reform (rapporteur: Marju Lauristin MEP, Estonia) that secured minimal protection for EU citizens including a right to be forgotten, "clear and affirmative consent" to the processing of private data by the person concerned and the obligation to explain the privacy policies in clear and understandable language.

Yet, more work is needed to make cultural big data trade more transparent and to develop best practice for the uses of big data in the service of cultural diversity, political pluralism, more effective and transparent knowledge exchange and for facilitating broader “reflexive modernisation”⁹, whereby risks of modern living are identified and solutions are co-invented . Dominant online service providers should not only become transparent regarding privacy terms and informing private individuals on the usage of their data but also with regard to informing society at large of the terms of their data trade and how this may affect the nature of the knowledge services provided to European citizens. Furthermore, good conduct for sharing the usage data between private and public institutions in the service of a more transparent and reflexive society – i.e. the ‘public good’ – should be developed and highlighted. The Council of Europe with its ‘soft regulation’ instruments would be well placed to lead such work.

9 Beck, Giddens & Lash, 1994

References

- Appadurai, Arjun. (2003). Archive and Aspiration. In J. Brouwer & A. Mulder (Eds.), *Information is Alive: Art and Theory on Archiving and Retrieving Data* (pp. 14-25). Rotterdam: NAI Publishers.
- Baron, Jaimie. (2014). *The Archive Effect: Found footage and the audiovisual experience history*. Oxon: Routledge.
- Beck, Ulrich, Giddens, Anthony, & Lash, Scott. (1994). *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*. Cambridge: Polity Press.
- Casalegno, Federico. (2004). Thought on the Convergence of Digital Media, Memory, and Social and Urban Spaces. *Space and Culture*, 7(3), 313-326.
- Couldry, Nick, & Powell, Alison. (2014). Big Data from the bottom up. *Big Data & Society*, 1(2). doi: 10.1177/2053951714539277
- Council of Europe. (2016). *Recommendation CM/Rec(2016)2 of the Committee of Ministers to member States on the Internet of citizens*. Strasbourg: Council of Europe Retrieved from [https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec\(2016\)2&Language=lanEnglish&Ver=original&Site=CM&BackColorInternet=C3C3C3&BackColorIntranet=EDB021&BackColorLogged=F5D383&direct=true](https://wcd.coe.int/ViewDoc.jsp?p=&Ref=CM/Rec(2016)2&Language=lanEnglish&Ver=original&Site=CM&BackColorInternet=C3C3C3&BackColorIntranet=EDB021&BackColorLogged=F5D383&direct=true).
- Council of the European Union. (2012). *Council conclusions on the digitisation and online accessibility of cultural material and digital preservation*. Brussels: Retrieved from http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/130120.pdf.
- Cusumano, Michael A. (2011). Platform wars come to social media. *Communications of the ACM*, 54(4), 31-33.
- David, Paul, & Greenstein, Shane. (1990). The economics of compatibility standards: An introduction to recent research. *Economics of innovation and new technology*, 1, 3-41.
- Ellis, John. (2012). Why Digitise Historical Television? *Journal of European Television History and Culture*, 1(1), 27-33.
- Ernst, Wolfgang. (2004). The Archive as a Metaphor. *Open*, 7, 43-46.
- Ernst, Wolfgang. (2013). *Digital Memory and the Archive* (J. Parikka Ed.). Minneapolis: University of Minnesota Press.
- Ernst, Wolfgang. (2015). *Stirrings in the Archives: Order from Disorder*. Lanham: Rowman & Littlefield Publishers.
- Foray, Dominique. (1995). Coalitions and committees: how users get involved in information technology (IT) standardization. In R. Mansell, R. Hawkins & J. Skea (Eds.), *Standards, Innovation and Competitiveness: The Politics and Economics of Standards in Natural and Technical Environments* (pp. 192-212). Aldershot, Brookfield: Edward Elgar.
- Hartley, John. (2012). *Digital Futures For Cultural and Media Studies*. Chichester: John Wiley & Sons.
- Hoskins, Andrew. (2011). Media, Memory, Metaphor: Remembering and the Connective Turn. *parallax*, 17(4), 19-31.
- Ibrus, Indrek. (2013a). Evolutionary dynamics of media convergence: Early mobile web and its standardisation at W3C. *Telematics and Informatics*, 30(2), 66-73. doi: <http://dx.doi.org/10.1016/j.tele.2012.04.004>
- Ibrus, Indrek. (2013b). Evolutionary dynamics of the mobile web. In J. Hartley, A. Bruns & J. Burgess (Eds.), *A Companion to New Media Dynamics*. London: Blackwell.
- Ibrus, Indrek. (2015). Histories of Ubiquitous Web Standardization. In A. Bechmann & S. Lomborg (Eds.), *The Ubiquitous Internet: User and Industry Perspectives*. London: Routledge.
- Katz, Michael L. , & Shapiro, Carl (1986). *Production compatibility choice in a market with technological progress*. *Oxford Economic Papers* 38, 146-65.
- Niggemann, Elisabeth, de Decker, Jacques, & Lévy, Maurice. (2011). The New Renaissance: Report of the 'Comité Des Sages' on bringing Europe's cultural heritage online (pp. 191). Luxembourg: European Commission.
- Puschmann, Cornelius, & Burgess, Jean. (2014). Big Data, Big Questions | Metaphors of Big Data. *International Journal of Communication*, 8, 1690–1709. doi: 1932–8036/20140005

- Reading, Anna. (2014). Seeing Red: A political economy of digital memory. *Media Culture & Society*, 36(6), 748-760.
- Torlasco, Domietta. (2013). *The Heretical Archive: Digital Memory at the End of Film*. Minneapolis: University of Minnesota Press.
- Uricchio, William. (2009). Moving beyond the artefact: Lessons from participatory culture. In M. van den Boomen, S. Lammes & A.-S. Lehmann (Eds.), *Digital Material: Tracing New Media in Everyday Life and Technology*. Amsterdam: Amsterdam University Press.
- Vonderau, Patrick. (2015). The Politics of Content Aggregation. *Television & New Media*, 16(8), 717-733. doi: 10.1177/1527476414554402
- Wactlar, Howard D. , & Christel, Michael G. (2002). Digital Video Archives: Managing Through Metadata. In *Building a National Strategy for Preservation: Issues in Digital Media Archiving*. Washington: Council on Library and Information Resources Washington, D.C. and Library of Congress.

Indrek Ibrus is Associate professor at Tallinn University's (TLU) Baltic Film, Media, Arts and Communication School (BFM). He is also the head and senior researcher at TLU's Center of Excellence in Media Innovation and Digital Culture (MEDIT). He has published extensively on media innovation, audiovisual media industries and regulation, mobile media, transmedia and cross-media production. He received his PhD from the London School of Economics and Political Science.