

NEW CINEMA HISTORY AND THE COMPUTATIONAL TURN

Deb Verhoeven¹

Abstract — This paper will outline how the digitisation of both the film industry and contemporary research practices bear on the work of the new cinema historian. How might the opportunities presented by an unprecedented proliferation of data for example, also challenge the unspoken assumptions and ordinary practices of conventional film studies research? And how might the ‘computational turn’ present opportunities (and challenges) for a revisionist cinema history at the intersection of qualitative historiographies (focussed on the social experience of the cinema) and quantitative research approaches such as data mining, empirical analysis and digital visualisations?

Index Terms — new cinema history, computational turn, digital humanities

As part of a broad disciplinary shift, from a focus on measuring the value of cultural artefacts to understanding the import of cultural flows, humanities researchers are increasingly turning to other disciplines to inform their research and to participate in the development of new approaches to the study of everyday life. For example, as cinema history extends its interest – beyond the histories of films alone – the kinds of information it engages with change also. This New Cinema History explicitly acknowledges the wider historical dimensions of everyday cinema experiences and its attendant industrial practices and involves a multitude of disciplinary approaches, including history, cultural geography, information management, computer science, geo-spatial science, social science, economics, statistics, and creative arts to name a few [3][4]. Underlying the New Cinema History is recognition that the cinema is neither economically, culturally nor socially isolated and that to fully explore its history cinema researchers must collaborate across disciplines, institutions and social locations.

If we understand the cinema as comprising institutional, social and commercial practices that are interdependent then it follows that new kinds of ‘evidence’, and new ways of organising this evidence, are necessary for our research. For example, much criticism has been directed at the way in which conventional film theory and film philosophy has typically characterised the cinema audience as a non-material or hypothetical entity, in stark contrast to the ‘presence’ or ‘evidence’ of film texts. On the other hand, for new cinema researchers there are pressing questions that arise when the focus of evidence collection is not on the film texts themselves but on the experiences and events that

occur around film consumption. The impact of this focal shift can be most clearly seen in the way new cinema historians have developed innovative digital research techniques in order to extend their scholarship in new directions and raise questions that weren’t previously possible. So instead of simply proposing that the release of film productions serve as an index of consumption, we might instead consider how the study of historically and spatially located consumption practices can reorient both our research questions and our research techniques.

Rather than measuring the comparative cultural value of film titles, the new cinema historian traces the flows and pace of historical change and measures the intensity of its dynamics; changes that may be intrinsic to industry practices and are just as likely not to be, and that are almost certainly only tacitly related to the properties of the film texts themselves. And whilst the availability of new research technologies has certainly had the effect of enabling these new approaches to the study of cinema, I am not suggesting that the computational turn in the humanities is the exclusive preserve of non-textual studies. Conventional film databases spiral around film texts and gather into their spin those aspects of a film’s production history and critical reception that give *a priori* gravity to the text’s centrifugal place. On the other hand, databases designed to capture information about cinema venues, require a different range and type of data, drawn from a wide array of sources, and which in turn propose challenges and risks for how new film historians conceptualise the very entities we seek to describe.

FILM HISTORY AND THE DATABASE

In the twenty-first century, not only are we all video store clerks (compiling vast personal film libraries of digitised cinema), we are also all information managers and archivists engaged in the constant organisation of our abundant collections, demonstrating our cultural mastery by uploading clips to file sharing sites, goading other collectors to appreciate our acquisitions and expertise through participation in various social media outlets. We make it easier for others to regard the breadth of our tastes by producing abbreviated mashups, proving our cultural credentials twice-over through our simultaneous blending of technical and curatorial flair.

For the serious digital film collector, these curatorial practices are best expressed in the form of lists. The popular practice of compiling and sharing film lists however also pinpoints some of the current limitations for how the rapid

¹ Deb Verhoeven, Professor and Chair of Media and Communication, School of Communication and Creative Arts, Deakin University, 221 Burwood Highway, Burwood Victoria 3125, Australia, deb.verhoeven@deakin.edu.au

digitization of film industries is currently made available to and understood by new generations of digitally skilled researchers. Locating films and arranging lists of the 'top' or 'best' or 'most' (and their converse, the 'worst') position films in a ranked ordering of similar entities (other films) rather than redefining the kind of cultural company that films might keep. For collectors and enthusiasts the value of their cinema collections lies ultimately in the pursuit of a perfected form of cultural connoisseurship in which film information interacts only with other film information. The implications of these delimited collection practices can similarly be seen in the information structure of online research resources such as the internet movie database (IMDb) which are, at their heart, like all databases, elaborated sets of lists.

The IMDb began in the late 1980s, the personal project of a film enthusiast, Col Needham who was particularly interested in capturing movie credit listings. Needham was soon joined by other enthusiasts who were likewise fascinated with developing and sharing registers of film actors. From modest beginnings, it expanded into one of the most comprehensive online film industry inventories containing information on more than 2 million film and TV titles and nearly 5 million industry members as well as boasting more than 30 million registered users.

There are evident limitations to the ways in which existing online film inventories, repositories and release catalogues such as IMDb manage the kinds of film information that might be of interest to the new cinema historian. 'Release' catalogues and datasets, for example, invariably preference the film rather than the place of release. Consequently they are frequently incomplete in relation to information about release location. You are unlikely to find movie release dates for Belize or Indonesia in the IMDb for instance. Furthermore these data sets are often retrospectively compiled and reliant on secondary information. As Hugh Amory notes, because their scope is often retroactive, these datasets bear the imposition of anachronistic "territorial and cultural inclusions and exclusions that were alien to their periods"[5].

Although they are valuable resources, these sorts of information databases are not specifically designed to answer key research questions that might be posed by new cinema historians, for example to calculate the relative size of distributor activity. Nor are they intentionally designed for the production of comparative cinema statistics or critical histories. Their data is often unavailable for export into more flexible analytical software and when it is possible to extract data for the purpose of detailed analysis this is typically a difficult process that requires some level of technical expertise. Accordingly, the primary users of filmographic catalogues are not cinema historians, information managers, analytical filmographers or cinema scholars, but members of the public, film buffs, students and so on who are content to navigate these databases using the small number of structured search fields provided.

Databases like the IMDb, with their broad aim for comprehension, are also prone to sweeping over the granularity of differences and distinctions that distinguish data derived from different sources. So seemingly straightforward data categories like 'Name' (Surname, First name) do not work well for capturing Chinese naming conventions for instance (a notable issue in IMDb in which the actor Chow Yun-Fat appears throughout the database as Yun-Fat Chow). And even though the film text is the basic organising principle for so much of their data, most film catalogues and indexes are based on a notion of the 'text' that ranges from the intentional to the ideal. Consequently they are unable to tell you very much about what was actually being watched, or what was available to see, and they can't represent the performance of a title in the context of a total number of screenings (including multiple formats and versions) for instance.

These types of catalogues then, carry only an incidental relevance to the pragmatic operations and concerns of the film industry, its investments and sales capacity for instance. For the new cinema historian on the other hand, it is crucially important to remember, how, why, by whom and for whom, films are manufactured and circulated, and to be mindful of the various specific uses to which a film is put. The changing contexts in which a film is viewed for example not only alter our understanding of the relative significance of the text but its very definition. The classic film title, *Ben Hur* exists in the IMDb in the form of multiple historical renditions each with their own separate entry; two silent versions in 1907 and 1925, the award winning 1959 release, a 2003 animation (also featuring Charlton Heston), and a TV series made in 2010. What the database doesn't describe is the sheer variety of different viewing experiences that even one of these iterations of *Ben Hur* inspired. For example, the Australian release of the 1959 title included a typical Run-Zone-Clearance distribution pattern across the country beginning in July 1960. But there was also a local, meticulously subtitled version of the film that screened at Greek language venues in the early 1960s, an Italian dubbed version imported by US distributors to Australia specifically for Italian diasporic audiences in urban and remote locations, various significant repertory seasons in the late 1960s, and an extended drive-in revival in the early 1970s. With these local variations in mind, we can see how the IMDb gives preferential treatment to the text as the instantiation of a production history rather than the outcome of distinct consumption histories. We can imagine what new insights and questions a consumption driven database might reveal, simply by organising its information around 'film engagements' rather than historically defined 'film productions' (see *The Cinema Context* [6] database or the *Cinemas and Audiences in Australia*[7] database for examples).

Not only do we need to develop and apply new methods for analysing the vast array of digital collections that now exist, we need to ensure our practices as new

cinema historians keep pace with developments in information design and management, data collection and research dissemination. There is for the moment a significant gap in cinema studies between catalogue sources (such as those managed by national archives and libraries) and scholarly research processes, and very little feedback between collection management and academic research for instance. This has prompted the development of a significant number of discipline specific database projects simply to ensure a closer connection between research analysis, evaluation and the collections they are based on.

BEYOND IMDB: NEW CINEMA HISTORY RESEARCH AND THE DIGITAL

Digital technologies have categorically changed the way we engage in the processes of cinema research. From the use of social media as a research tool, to communication technologies that cement collaborative activities (email, Skype, near field technologies), the meaning of cinema research, its workflows and outcomes have been fundamentally transformed.

Our use of the internet as a research tool is now so ingrained as to be effectively ubiquitous. Search engines, in particular Google, are becoming increasingly critical as more and more of the world's resources are digitized and made accessible online. 'Findability' and 'searchability' are the contemporary measure of successful research design, central to a new cache of research techniques such as data clustering, data mining, visualisation and the application of algorithms to name only a few, all made possible by the unprecedented abundance of information at our fingertips.

Increasingly, across all academic disciplines, including screen studies (and perhaps especially given the vast volume of data we generate through moving image files), we are challenged to manage and understand an overflow of data. As Dan Cohen has noted, this is in clear contrast to the challenges of the past, in which historians were confronted with, and developed methodologies for addressing, a defining absence of information:

It is now quite clear that historians will have to grapple with abundance, not scarcity. Several million books have been digitized...and nearly every day we are confronted with a new digital historical resource of almost unimaginable size. [8]

For new cinema historians the data challenges are particularly acute. By expanding the object and scope of cinema research beyond the film text itself, we make relevant a wider range of information types, formats and sources. So the question for the new cinema historian is actually much more complicated because it is not just the traditional objects, in this case the films, which we are interested in. By expanding the range and type of information that is relevant to our study (government reports, ordinances, building or police records, regulatory legislation, tax files, oral histories, marketing materials,

industry archives, maps, box-office data, phone books, ticket stubs, newspaper advertisements just to name a small few) we correspondingly expand the amount of information available to us and lift the significance of our ability to locate, collect, aggregate, curate, manipulate and analyse different data formats from different sources and for which available tools are proving increasingly inadequate.

Stephen Ramsay for example identifies the shortcomings of existing tools provided by Google for navigating its vast online library holdings:

As a search tool, Google is hard to beat. By providing lookup access to the contents of the books, it provides a facility that no library has ever been able to offer in the history of the world. Yet as a browsing tool – as a tool for serendipitous engagement – it falls far behind even the most rudimentary library. It can successfully present books on gardening, but because all categorization within Google Books is ultimately a function of search, it has a hard time getting you from gardening to creation myths, from creation myths to Wagner, and from Wagner to Zappa. It may sound perverse to say it, but Google Books (and indeed, most things like it) are simply terrible at browsing. The thing they manage to get right (search) is, regrettably, the one thing that is least likely to turn up something not already prescribed by your existing network of associations. In the end, you're left with a landscape in which the wheel ruts of your roaming intellect are increasingly deepened by habit, training, and preconception. Seek and you shall find. Unfortunately, you probably won't find much else. [9]

Ramsay recognises the value of serendipitous discovery in humanities research as crucially missing from the current suite of research tools. How we restore serendipity to our digital research practices, at the time of writing, remains an open question. Jon Orwant, speaking as the engineering manager at Google, has expressed the company's intention to find algorithms that can accommodate not just personal research preferences but collective ones (like a shared bookshelf), and which are based on more sophisticated analytics with multiple classifications in order to increase what he calls "deep" serendipity [10].

The contemporary cultural researcher's situation is made more complex still by a multiplying number of analytic tools and methodological options. At the other end of the research workflow from 'search', lie further critical challenges. We need to be able to manipulate data – both quantitative and qualitative. At best, most specifically developed cultural databases provide services that enable sophisticated ways of searching digital object collections and the descriptive metadata assembled by curatorial and institutional experts. But they rarely provide the research processes that would enable the researcher to exploit these sources.

Furthermore, as Toby Burrows notes, humanities researchers produce and make use of other kinds of evidence which don't fit neatly into the binary of qualitative and quantitative data. And within the disciplines there is confusion about what 'data' itself might be. For Burrows this is a particularly acute issue in the humanities in which the distinction between 'data' and 'sources of data' – or between evidence and sources of evidence – is frequently blurred:

It would be analogous to describing the stars and galaxies as an astronomer's 'data' when, in fact, the actual physical objects are clearly distinguishable from the observations relating to them—and these observations form the data which the researcher uses and analyses. The difficulty for the humanities is that they do not deal exclusively with physical phenomena. They are also concerned with more abstract entities like texts and works, which are conceptual entities as well as their physical manifestations. [11]

Extending Burrows' point, we can see how for many text-focused film researchers, digital film texts might appear as if they *are* 'data' rather than as a container *of* data. But the digitisation of films should not be understood as either the equivalent of, or indeed a replacement for, data-centric research. The new cinema history's redirection, away from the discipline's previous focus on the film text, is inadvertently helpful in elaborating this distinction.

Burrows identifies several specific working challenges which hinder data-driven research for digital humanities scholars such as new cinema historians:

- It is difficult to define 'data' in the humanities in a consistent (i.e., machine-processable) way;
- It is difficult to identify and model generic research processes, since research methods tend to be poorly documented and little discussed, or are regarded as matters of common sense;
- There has been a strong tendency towards project-specific digital solutions which cannot be aggregated into a more general e-research framework;
- It is difficult to separate analysis and research outcomes from the source materials—one researcher's publications quickly become another researcher's evidence or data;
- There is a gulf between the research processes of academic researchers and the curatorial processes of the cultural institutions which hold most of the source materials [11]

Additionally, the size and scope of newly abundant data is both an enormous opportunity and an equal challenge. For present-day and future cinema historians, undertaking research informed by empirical approaches to everyday social behaviour will rest on an ability to store, aggregate and combine vast amounts of data and then use the results to perform deep analyses in order to glean insight. It is estimated that in 2010 more than 4 billion people (around 60% of the world's population) owned smart phones. MGI estimates that commercial companies globally

stored more than 7 exabytes of new data on disk drives in 2010, while consumers stored more than 6 exabytes of new data on devices such as PCs and notebooks. One exabyte of data is the equivalent of more than 4,000 times the information stored in the US Library of Congress [12]. Big data is not only changing the way we approach cinema research, it will inevitably make inroads on industry practice: supporting commercial decision making with automated algorithms (such as around distribution windows and film itineraries for instance) or enabling even more specific market segmentation and product innovation.

For researchers there are also benefits. Analysing spatial and temporal distribution patterns (looking at screenings data but also mapped against transport analytics), analysing attendance patterns (by combining multiple datasets such as ticket sales histories, weather predictions, and seasonal attendance cycles for example), aggregating government datasets, analysing box-office data, identifying audience segments (down to 'micro-segments'), producing audience behaviour analysis (drawing on real-time location data from mobile phones, automotive telematics or image data from retail centre video cameras), and sentiment analysis (monitoring social media applications). The analysis of large datasets enables researchers to move beyond linear approximation models to complex models of greater sophistication. New cinema historians might examine the precise impact of several variables on box-office performance or cinema attendance for example. Using data-mining techniques cinema researchers can better understand and predict peak cinema consumption patterns on a global basis. Large datasets also allow researchers to identify and interrogate 'rare events' and 'low incidence populations' that could otherwise escape detection or fall below a threshold of reliable statistical significance in a smaller sample analysis. Conversely, the analysis of small cinema markets or audience segments in context becomes more feasible and accurate using these techniques.

NEW CINEMA HISTORY AS A 'POST-DISCIPLINE'.

From the outset the new cinema historian accepts that we produce only one type of knowledge amongst many, and that others will use the knowledge we produce in ways we may not have anticipated. New cinema historians, working across different methodologies, recognise that we can accomplish more collectively than we can as individuals.

Because our techniques and technologies necessarily draw from a variety of fields including statistics, information management, geospatial science, computer science, applied mathematics, and economics we need to adopt a flexible, contingent, cross-disciplinary approach to answering questions of the available data. We also need to become comfortable with different 'registers' of approach, combining academic and commercial methodologies for instance, or scholarly and amateur expertise. In galvanizing research teams made up of differently specialized

researchers that are *problem* oriented, rather than discipline or program specific, we propose a model for collaboration that is itself mindful of the temporalities of contemporary academic practice. In the words of Mario Biagioli, this model foregrounds, “a new and distinct pattern of postdisciplinarity”[13]. For Biagioli modular research practices constitute:

neither a family of disciplines, nor a new bud or branch of the tree of knowledge. It is a problem specific collaboration that takes place within a limited temporal window and in places that may have little to do with standard departments and institutes... What matters the most is to maximize the quality of one’s skills and to expand their range so as to be able to move from one fruitful collaboration to the next. [13]

By working towards collective research objectives, the new cinema historian challenges entrenched assumptions about the value of the lone-wolf scholar-author. This aspiration for interconnectedness might also extend to the ways we publish and distribute our research as well. If we understand our work as belonging to, and in the service of, different communities (our own and the ones we are researching) then we might measure the value of our research in terms of community engagement rather than the familiar academic metrics of comparatively ranked output.

In keeping with our cross-disciplinary curiosity the new cinema historian is mindful of the position of our data amongst a collection of collections. The value of a collection in the digital space its not its own size and scope but what it can contribute to other national and international collections. With the ability to find and describe new pathways through an ever-proliferating archive of data, technologies, research tools and resources we can collaboratively answer larger problems across disciplinary boundaries, enable new discoveries from previous data, create longitudinal time-series analyses and avoid duplication of effort. The future for a renewed cinema scholarship then, lies in organizing how we collect, arrange, describe and share information; in better understanding its context and managing its evolution and use.

CONCLUSION

The types of information that were once regarded as the distinctive preserve of the sciences (including the social sciences) are now an everyday part of the way we experience the world. The quantitative and qualitative information that resulted from carefully crafted research practices, typically prepared and performed by specialised researchers, are now widely available to anyone. This ready availability of data, and the ability to interrogate it in pursuit of research questions for which it may not have been intended, produces interesting challenges for new cinema historians.

To date our data requirements have been relatively small (although mass digitisation projects, news feeds and increasing volumes of accessible sound and media, as well

as censuses and social surveys are pushing the boundaries and making new opportunities available). On the other hand, our data is characterized by being highly complex, heterogeneous and interlinked. The real value of our information lies between various objects of study, not within them and requires collaboration (including crowdsourcing) and the building of new and extended research capacities and connections.

The new cinema history describes the intersection between a revised qualitative cinema historiography (focused around an industrially informed and consumption attentive view of the cinema) and the use of innovative information systems (inspired by new research approaches such as data mining, empirical analysis and digital visualisations). The cross-disciplinary requirements and participatory technologies of digital research provide unique opportunities to new cinema historians since it is precisely through the opening up of collaborative research and publication partnerships that we move beyond the tautologies of historical narratives fixated by the film text.

ACKNOWLEDGMENTS

I would like to acknowledge the feedback and contributions of my colleagues Colin Arrowsmith, Richard Maltby, Kate Bowles and Mike Walsh in the preparation of this paper.

REFERENCES

- [3] Maltby, R, “New Cinema Histories” in R. Maltby, D. Biltereyst, & P. Meers (eds) *Explorations in New Cinema History: Approaches and Case Studies*, Wiley-Blackwell, 2011, pp. 3-40
- [4] Maltby, R, Bowles, K, Verhoeven, D, & Walsh, M. *New Cinema History: A Guide for Researchers*, Wiley-Blackwell, 2011.
- [5] Amory, H, “Pseudodoxia Bibliographica, or When is a Book Not a Book? When It’s a Record” In *The Scholar & the Database: Papers Presented on 4 November 1999 at the CERL Conference Hosted by the Royal Library*, Brussels, ed. Lotta Hellinga, Vol. 2, 2001, p.4.
- [6] Dibbets, K, (ed), *Cinema Contexts* [Website] <http://www.cinemacontext.nl>,
- [7] Verhoeven, D, (ed), *CAARP* [Website], <http://caarp.flinders.edu.au/>
- [8] Cohen, D ‘Interchange: The Promise of Digital History’, *The Journal of American History*, Vol. 95, No. 2, Sept. 2008, <http://www.journalofamericanhistory.org/issues/952/interchange/index.html>
- [9] Stephen Ramsay, *The Hermeneutics of Screwing Around; or What You Do with a Million Books*, 2010, <http://lenz.unl.edu/>
- [10] Weinberger, D “[2b2k] Jon Orwant of Google Books”, *Joho the Blog*, March 29, 2010, <http://www.hyperorg.com/blogger/tag/google-books/>
- [11] Burrows, T, “Sharing Humanities Data for eResearch: conceptual and technical issues” in N. Thieberger et al (eds) *Sustainable Data from Digital Research: Humanities perspectives on digital scholarship*, University of Melbourne, Melbourne, 2011, p. 5
- [12] Manyika, J, Chui, M, Brown, B, Bughin, J, Dobbs, R, et al, *Big Data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute, 2011, p.3.
- [13] Biagioli, M, ‘Postdisciplinary Liaisons: Science Studies and the Humanities’, *Critical Inquiry*, Vol. 35, No. 4, 2009, p. 821.

CITATION: Verhoeven, D. (2012) ‘New Cinema History and the Computational Turn’, *Beyond Art, Beyond Humanities, Beyond Technology: A New Creativity*, Proceedings of the World Congress of Communication and the Arts Conference, University of Minho, Portugal