

Archives and Identity in the Context of Social Media and Algorithmic Analytics: Towards an Understanding of iArchive and Predictive Retention

(Version accepted for publication, November 24, 2017)

Holger Pöttsch, UiT Tromsø

Abstract

The present article reconceptualises the archive in the context of digital media ecologies. Drawing upon archival theory and critical approaches to the political economy of the Internet, I account for new dynamics and implications afforded by digital archives. Operating at both a user-controlled explicit and a state- and corporate-owned implicit level, the digital archive at once facilitates empowerment and enables unprecedented forms of management and control. Connecting the politics and economy of digital media with issues of identity formation and curation on social networking sites, I coin the terms iArchive and predictive retention to highlight how recent technological advances provide both new means for self-expression, mobilisation, and resistance, and afford an almost ubiquitous tracking, profiling, and, indeed moulding of emergent subjectivities.

Keywords

Archive, networks, algorithmic analytics, big data, surveillance, social media, identity, prediction, data doubles, digital technologies

Introduction

In his book *Reverse Engineering Social Media*, Robert Gehl (2014) takes up the recent emergence of so-called socialbots - algorithmically designed and maintained profiles on social media platforms that actively befriend and interact with human users. According to him, the

apparent success of these programmes, that are increasingly taken for real human users, does not reflect impressive advances in artificial intelligence. Rather, Gehl argues, “socialbots are a reflection of our activities within social media; for the machines to work, we ourselves have to be trained to be machinelike” (23). As such, instead of making machines more human, contemporary digital technologies might just as well entail a gradual deterioration of what it means to be human. The present article will interrogate such potential performative effects with focus on social media and digital archives.

Drawing upon advances in archival research and critical studies of the political economy of the Internet, I will ask the question what happens once the archive not any longer only looks back in time managing access to traces and documents informing what we remember both collectively and individually, but takes part in forming the future by systematically offering nudges and pokes based on user-profiles that turn mere possibilities for future conduct into probabilities and even actualities. During this inquiry, I coin the term iArchive to account for the ambivalent affordances of corporate-owned social media applications and argue that the present conjunction of increasingly ubiquitous digital networks, largely automated practices of surveillance and user-profiling, and algorithmically tailored feedback loops creates a predictive form of retention that not only aids the curation of individual pasts, but also - and increasingly - creates the subjectivities and performances that actively shape the future.

Theories of the archive

Marlene Manoff (2004) starts her overview over *Theories of the Archive from Across the Disciplines* with the observation that ‘[m]ost writers exploring the concept share a notion of the archive as a repository and collection of artifacts’ (10). According to her, the archive comprises a variety of institutions ranging from libraries and museums to ‘the entire extant historical record’

(10). Increasingly, these institutions rely upon digital networks and databases of varying sizes and catchments to fulfil their designated roles.

Indeed, the archive is a much-discussed term. As for instance Derrida (1995: 1-2) observes in his Freudian exploration of what he terms *Archive Fever*, deriving from Greek *Arkhe*, the term archive denotes both authority and origin, and therefore from the very beginning closely connects with both institutional (state) power and history. From this point of departure many scholars have highlighted various aspects of this ambiguity leading up to a series of works critiquing the often-tacit power over shared pasts wielded by archives and archivists. Greetham (1999), for instance, alerts to the significance of implicit frames for selection and retrieval underlying allegedly neutral and comprehensive archival collections, arguing for the fact that archival practices at a fundamental level pose questions of the ‘social formation of agency’ (3). Following a similar critical trajectory, scholars such as Richards (1993), Stoler (2009), Carter (2006), and Haebich (2016) have directed attention to the close relation between archives and colonial and imperial power.

Richards (1993) for instance elaborates upon the practices of knowledge production under British colonial rule that often remained detached from the factual life worlds of colonial subjects but that, nevertheless, gave rise to archives with the power to frame and predispose the cultures, politics, and individual subjectivities of entire continents. Criticising such power-laden practices of knowledge production and management, Richards writes about the ‘fantasy of the imperial archive’ (6) that only retained a loose connection to lived realities in the colonies. Arguing in a similar direction, Carter (2006) has alerted to the significance of attending to silences and omissions when assessing the powers of archives and archiving, while Haebich (2016) points to the importance of living cultural heritage as archives of indigenous populations that counter the

often-oppressive functions of state repositories. Studies such as the ones referenced above bring forth archives as sites from which (state) power derives its legitimacy. These same archives, however, also enable challenges to, and democratic redistributions of, this archontic authority.

Moving from (post-)colonial archives to liberal ones, Joyce (1999) shows how a gradual transformation of archives from secluded repositories accessible by elites to increasingly open institutions providing public services, aided the constitution of a liberal citizenship. Somewhat conflating the functions of libraries, museums, and archives, Joyce shows how the same institutions that manage colonial power/knowledge configurations (as critiqued by for instance Richards 1993) gradually opened up and became vital for processes of democratisation and the formation of a liberal political order in Britain. This significance of a democratic control over archives is already present in Derrida's (1995) thinking when he writes in a footnote that '[e]ffective democratization can always be measured by this essential criterion: the participation in and the access to the archive, its constitution, and its interpretation' (4; note 1).

Osborne (1999), too, directs attention to an archive-politics nexus. Redirecting focus from apparently static institutions to the everyday practices of what he terms 'agents of the archive' (52), Osborne asserts a contingency of archival functions upon the societal frames within which they operate. According to him, the role of archives in specific periods of history can be seen 'as symptoms of some of the leading characteristics of [...] society' at a given time (52). As such, archives apparently both reflect and refract received societal orders and frames.

Reiterating this double-notion of the archive as both historical agent and symptom, Lynch (1999) pairs Derrida's (1995) 'archontic power' (3) comprising among other things the 'hermeneutic right and competence [...] to interpret the archive' and make the deposited documents 'call on or

impose the law' (2), with a different, more subversive, form of archival agency. Identifying a 'dialectics of archontic and anarchic power', Lynch (1999: 71) writes:

What Derrida calls 'archontic power' - control over the authorship, collection and interpretation of a body of writings - was supplemented, and at times counteracted, by what might be called 'anarchic' power - a resistive power characterized by control over the drafting, destruction and dissolution of records to enhance the equivocality of interpretations and accusations.

Lynch then moves on to illustrate this anarchic power of archives and archival practices with reference to the role various forms of documentation and document destruction played in the official investigation of the Iran-Contra affair conducted by US authorities in the late 1980s.

After having identified a potential of archives to balance and counter-act state power rather than merely reinforcing it, Lynch (1999) moves on to problematize the very 'archontic infrastructure' (79) that predisposes the assembling, storage, accessibility, and dissemination of documents with assumed value for particular communities. Adding the dimension of technology to this infrastructural inquiry, he observes that '[t]he recent proliferation of electronic means for reproducing and disseminating documents and entire archives has begun to disrupt the traditional exclusiveness of scholarly access' (75) effectively creating a 'popular archive subjected to mass visitation, reproduction and dissemination' (75-76). Writing in the late 1990s, Lynch's words seem to presage the archival functions of ubiquitously networked mobile media and peer-to-peer computing that enabled the disruptive operations of distributed digital counter-archives such as those curated and released by *WikiLeaks* (Assange, 2015; Harrison, 2015),ⁱ but that also aided the proliferation of propaganda through what might be termed fake archives (Herrman, 2016; Greenwald and Klein, 2016).

Lynch's (1999) work moves the present article toward its main theme - the impact of contemporary digital network technologies on forms and functions of archives that lead up to the emergence of the digital archive. This archive-technology nexus has been investigated earlier, and, once again, we can start with recourse to Derrida's *Archive Fever* (1995). Taking a rather techno-determinist stance, Derrida asserts that 'the technical structure of the *archiving* archive also determines the structure of the *archivable* content [...] The archivization produces as much as it records the event' (17; emphasis in original). Arguing that psychoanalysis would not have emerged in its present form if Freud had used email rather than handwritten letters, Derrida then concludes that 'electronic mail [...] is on the way to transforming the entire public and private space of humanity, and first of all the limit between the private, the secret (private or public), and the public or phenomenal' (17).

In a series of studies from the early 2000s onwards, Wolfgang Ernst (2003, 2007, 2013) has worked through the relations between technical media and archival institutions and practices. He analyses the technical conditions under which archives operate and that predispose their functions and effects - from classical ones that order written documents through analogue filing systems, via challenges to the alphanumeric logic of such traditional archives posed by 20th century audio-visual records, to the deterritorializing dynamics of contemporary digital networks. In sum, Ernst (2003) argues that algorithm-driven digital archives are based more on counting (zählen) than recounting (erzählen), as such instigating a post-narrative approach to the past in line with Manovich's (2001) 'database logic' (218) of contemporary culture that presents the world 'as an endless and unstructured collection of images, texts, and other data records' (219) open for navigation and search, but somewhat resistant to linear storytelling.

For Ernst (2013), the digital archives forming the contemporary Internet function as searchable databases that enable smooth and ubiquitous access to, and transfer of, knowledge, documents, and information. He writes that the silent, 'source-oriented' and 'file-oriented' storage sites of modernity are transformed into the current use-oriented and constantly evolving 'dynarchive' (82):

The notion of the archive in Internet communication tends to move the archive toward an economy of circulation: permanent transformations and updating. [...] Repositories are no longer final destinations but turn into frequently accessed sites. Archives become cybernetic systems. The aesthetics of fixed order is being replaced by permanent reconfigurability. (99)

In particular, the dimension of cybernetic reconfigurability - a constant adaptation of archival forms and functions in line with automated assessments and feedback loops - becomes crucial for an understanding of the new dynamics and tendencies afforded by digital archives in general and by what I, in the present article, term iArchive and predictive retention in particular.

Ernst's approach is based on a media archaeological method that moves attention from semiotics and meaning contained in documents to the technical apparatus that predisposes the modus of operation of these documents. While enabling innovative insights that often remain outside the scope of traditional approaches in the humanities, Ernst's method exhibits a tendency of bracketing the historical, socio-economic, and political embeddedness of technology. Jussi Parikka (2013), for instance, notes that

[i]t is striking how quickly Ernst moves away from even hinting at any sociohistorical contexts for such [digital] devices, emphasizing the objects in themselves - again a demonstration of his cool object focus, which by way of methodological choice saves itself, too, from discussing messy politics of technology. (13)

According to Parikka, what remains lacking in Ernst's work is 'a distinct political emphasis that would be able to talk about the political economy of increasingly closed [...] and black-boxed media technologies' (14). Not least this factor also limits the potential impact of practices such as DIY solutions, reverse engineering, or circuit bending advocated by Ernst (2013) as possible means to resist and re-appropriate technologies including digital ones.

So far, the present article has described approaches to archives in general subsequently homing in on Ernst's dynarchive. The latter term describes technical characteristics of digital archives that enable dynamic adaptations through cybernetic feedback loops, constant change and updating, as well as ubiquitous accessibility, and that draw attention to the distinct micro-temporalities and non-narrative logics specific to digital machines.

To differentiate the term 'dynarchive' further, Ernst (2013) suggests a series of additional concepts that elaborate upon specific technological dynamics. He proposes 'latent archive' (82) to account for the preservation of algorithmic or digital art that cannot be achieved through storage of specific objects, but depends on a source code through which an, in principle, infinite variety of contingent art objects can be brought to emerge. In a similar manner, Ernst describes the contemporary Internet as a 'transarchive' (84) or 'generative archive' (84) to highlight that this technology, in contrast to classical file-based archives, exhibits a 'mathematical [rather] than classificatory topology' (84) that subsumes various medial forms under the binary logic of digital code 'render[ing] commensurate texts, images, and sounds' (85) and that constantly produces new content and configurations. Lastly, Ernst uses the term 'streaming archive' (87) to describe the discrete temporalities of digital data that entail new dynamics of acceleration and continuous access.

In spite of their usefulness in accounting for specific technical dynamics, none of Ernst's terms adequately captures the relations between contemporary archival technologies and the political economy of digital capitalism including environmental and societal impacts. Nor do these terms explicitly address new dynamics of identity curation and creation through digital self-archiving on social networking sites. Following up on both Parikka's criticism and Ernst's terminological innovations, I propose the terms iArchive and predictive retention to alleviate such blind spots in established terminologies.

The politics of digital archives

Digital network technologies have had profound impacts on most areas of human life including archival institutions and practices. The quality of the relations between technologies and individual as well as collective practices (and, therefore, politics), however, is far from clear-cut. As for instance Pötzsch (2017) has argued, rather than speaking about determination and monolithic effects and impacts, it appears wise to perceive of technology, politics, society, and individuals as closely intertwined, mutually dependent, and as constituting one another in constant processes of exchange. In a similar manner, Dafoe (2015) has pointed out that '[t]he question should not be [...] whether technological determinism is right or wrong, but a set of questions of degree, scope, and context' (1050) that enable an understanding of technology's manifold and ambiguous functions and effects, and that move into view its varying degrees of autonomy and power. Regarding the archive, this form of pragmatism enables a productive piecemeal approach to the specific politics of particular archival technologies and their distinct contexts and practices of use.

One such context is highlighted by Cook (2013) who proposes a four-stage model to understand the development of archival practices and institutions from pre-modernity, via modernity to post-

modernity. Somewhat de-emphasizing the ultimately ambiguous affordances of digital networks and media (in particular 97 and 113-116), he asserts that the fourth archival paradigm is characterized by a beneficial role played by new digital technologies of the archive with regard to empowerment and agency. He writes:

[A]rchives as concept, as practice, as institution, and as profession may be transformed to flourish in our digital era, especially one where citizens have a new agency and a new voice, and where they leave through digital social media all kinds of new and potentially exciting, and potentially archival, traces of human life, of what it means to be human. (97)

Cook is partly right when identifying certain potentials for empowerment and self-expression connected to digital technologies. Authors such as Handley (2013), Assange (2015), and Harrison (2015), among others, would probably agree that digital networks and in particular encrypted and anonymized peer-to-peer interaction has had a beneficial impact upon movements and initiatives attempting to exert some control over, resist, challenge, and ultimately subvert the received archontic power wielded by state- and non-state actors. On the other hand, scholars such as Gehl (2014), Harcourt (2015), Fuchs (2012, 2017), Andrejevic (2007, 2013), or Pöttsch (2015a&b) have taken a more critical stance toward the ultimately ambiguous affordances of apparently progressive new technologies.

In contemporary digital media ecologies, a handful of global businesses controls and commodifies the majority of online traffic and data storage (Fuchs, 2017; Lanchester, 2017). Multinational corporations behind such applications as *Facebook*, *Twitter*, or *Google* combine social networking, cloud storage, online search, and electronic communications. In all these cases, the offered services and products are apparently free as no or only very modest fees are charged

by providers. Nevertheless, these companies exhibit impressive profit rates that rapidly transform small upstarts into vast corporations with global reach.

Fuchs (2012) has described and criticised the commodification strategies of digital capitalism enabling such profit rates. Writing about the role of the ‘internet prosumer commodity’ (146), he states:

A widely-used accumulation strategy is to give the users free access to services and platforms, let them produce content, and to accumulate a mass of prosumers that are sold as a commodity to advertisers. No product is sold to users; the users are sold as a commodity to advertisers. (144)

To make this work in economic terms, two different forms of the digital archive need to be distinguished. On the one hand, an overt and explicit surface archive that to a certain degree remains under the control of the individual prosumer, and a tacit or implicit deep archive, controlled by corporate interests, that operates underneath the radar of average users on the other.

Questions regarding access to and control over the deep archives storing user data and metadata are not only an issue of economics and corporate interests. As among others the revelations by Edward Snowdenⁱⁱ have shown, also secret services and other clandestinely operating state institutions have a vested interest in the constantly evolving databases of the deep layers of digital archives. Scholars such as Andrejevic (2013), Hogan (2015c), Lyon (2014), Scahill and Greenwald (2016), and Pöttsch (2015b & 2017) have addressed such political dimensions of data storage, processing, and retrieval in contemporary new media ecologies.

What becomes conceivable, then, is a digital archive that emerges from, and continuously evolves through, implicit background processes, and that harvests, on behalf of both private and state interests, the mundane daily activities carried out by users on digital networks. Gehl (2014) offers modern computers' von Neumann architecture as an analogy to illustrate this relation between an implicit and an explicit dimension of digital archives in the context of social networking sites.

According to Gehl (2014: 41-70), von Neumann differentiates computation into real-time processing and memory functions that are carried out by Central Processing Units (CPUs) and Random-Access Memory (RAM), respectively. Drawing upon the examples of *Facebook*, *Instagram*, and *Twitter*, Gehl argues that these applications in essence only provide empty templates that users then fill with self-generated content. In his analogy, users operate as CPUs that process data in real-time, but lack the capacity for storage and retrieval. Corporate- and state-owned implicit deep archives, on the other hand, resemble RAM that retains the processed data and controls access for further computation. This model aptly illustrates that digital technologies only apparently, or only to a certain degree, empower networked individuals and, in reality, cede enormous amounts of knowledge and power to multi-national corporations and, by extension, to state agencies.

Even though Gehl provides an accurate description of the surface-depths dynamics of contemporary digital networks, one aspect is missing from his inquiry. A palpable material dimension largely remains outside the purview of his approach that de-emphasises environmental issues, relations of production, and physical infrastructure. However, as for instance Hogan (2015a&c) has argued, social media archives are dependent upon a vast infrastructural 'underbelly', the costs and implications of which are rarely publicly discussed. According to her, *Facebook* obscures number, location, size, and energy consumption of the data centres upholding

its activities, this way blinding us ‘to the potential environmental costs of our everyday obsession with self-archiving’ (2015a: 5). In a similar manner, she connects the NSA’s Utah data centre, where US authorities store and assess intercepted global data flows, to local politics and environmental concerns thus problematizing issues of agency in complex socio-technical networks (2015c).

As a solution, Hogan (2015b) proposes an approach based on new materialism that allows us to respond to ‘dominant discourses and conceptual frameworks’ that hide these factors from view. Coining the metaphor of digital archives as ‘dumpsters’ (16), she urges us to pit ‘the archive’s orderly ambition up against the dumpster’s stinking mess’ (8) to make us take seriously the physical consequences of apparently clean and empowering digital technologies.

What has been said so far brings two different, yet closely related, dimensions of digital (dyn)archives into view. Firstly, an overt and largely user-controlled explicit ‘surface’ archive documenting and disseminating online individual memories, daily experiences, and personal expressions, and secondly, an implicit ‘deep’ archive that is unwittingly produced by users in and through their daily interactions with this surface archive and other networked environments. While explicit archives enable a limited form of user agency and conscious self-expression, implicit ones are assembled, owned, controlled, and instrumentalized by multinational companies and state actors, and largely remain outside the sphere of influence of individual users. The dialectics between the two constitutes a core dynamic of contemporary digital capitalism that constantly oscillates between ‘the poles of control-freedom’ (Chun, 2006: 6). The following section will develop the terms iArchive and predictive retention to focus the discussion on such implicit and explicit dynamics of identity management and curation on corporate social media.

iArchive and Predictive Retention: Identity Curation and Management on Social Media

Similar to other digital technologies, digital archives have a series of apparently contradictory characteristics. They emerge as at once material (located on physical servers) and de-territorialized (mobile and almost ubiquitously accessible), as both tacitly tailored to individual preferences (algorithmically customizing data streams in correspondence with developed profiles) and as abstractly massifying (enabling population-level pattern-of-life analysis), as dependent upon human input (in form of unpaid user labour) and increasingly autonomous (automatically acquiring, sorting, and instrumentalizing user data), as empowering (providing opportunities for expression, organisation, and mobilisation) as well as oppressive (facilitating unprecedented forms of surveillance, micro-management, and control), distributed among peers (through IP/TCP protocols) as well as centralized (in form of DNS and physical hubs), and as intimate (enabling the curating of miniscule details of personal lives) as well as intimidating (prone to peer-pressures and tacitly accessible by institutions and other users).

This ambiguity in forms and functions of digital technologies retains its relevance for the concept of iArchive that homes in issues of digital archives on social media practices and dynamics.ⁱⁱⁱ The ‘hip tricky little ‘i’ (Andrejevic, 2007: 4) points to ‘a timely double meaning, both solipsistic customization and the democratic promise of the ability to talk back’ (5). As such, in correspondence with such critical uses of the same prefix in terms such as iSpy, iCulture, iManagement and iMedia (Andrejevic, 2007), iWar (Pöttsch, 2015a), iBorder (Pöttsch, 2015b), and iSlave (Qiu, 2016), the ‘i’ in iArchive refers to the fact that the celebration of a ‘so-called interactive revolution [...] remains both premature and largely unexamined’ (Andrejevic, 2007: 5).

Gesturing towards these critical approaches, the term iArchive brings together explicit forms and practices of self-curating and self-expression on social networking sites with the implicit surveillance/management regimes that enable key actors to capitalize upon such activities. As a special form of the digital archive, iArchive offers critical perspectives on individual appropriations as well as the political economy and the socio-political implications of surface-versus-depth dynamics of retention in social media. The following sections will illustrate this further.

Designing subjects: Surface dynamics of iArchive

At an explicit user-driven level, digital technologies in general, and social media in particular, offer new means of self-presentation and expression. As van Dijck (2007) has shown, technologies such as digital photography, life logging, and blogging provide new means for performing and negotiating personal and collective identities. Identifying ‘digitization, multimediatization and googlization’ (150) as key dynamics of change, she details how computers impact upon the management of memories enabling new forms of self-curation at an individual and collective level. Making a similar argument, Walker Rettberg (2014) details the role played by blogs, selfies, lifelogs, activity trackers, and various instruments of quantification in user-led practices of identity management in new media ecologies.

Drawing attention to social media, Saker and Evans (2016) and Saker (2017) have investigated how location-based social networking sites such as *Foursquare* mesh online and offline practices leading to new forms of self-presentation and self-curation in digital environments. Saker (2017) identifies strategies employed by users to capitalize upon (in identity terms) automatic tracking and check-in functions that constantly share locations with a network of peers, and argues that social networking technologies enable ‘new front stage ways for people to present themselves to

others' (935). In a similar study, Saker and Evans (2016) address how users consciously relate to, and exploit for their own purposes, the implicit archival functions of social networking applications, as such, pointing to concrete practices of active (re)appropriation. Both studies thus reiterate Gillespie's (2014) argument about a 'a multidimensional "entanglement" between algorithms put into practice and the social tactics of users who take them up' (183) and home in these issues on users' engagements with code-based affordances of social media.

Other studies have focused on limitations and potentials for abuse built into digital self-curating technologies. Van Dijck's (2014) notion of 'dataism', for instance, describes drives towards ubiquitous (self-)quantification as inherently ideological and conducive to neoliberalism, while Drucker's (2010) concept of 'capta' (as an alternative to data) implies that something is not freely given by users (data), but captured by commercial and state actors (capta). Walker Rettberg (2014: 68-69) quotes both scholars when arguing that digital media enable both self-expressive empowerment and new forms of subjectification, micro-management, and control.

In a similar manner, Bucher (2012) has shown how *Facebook's* edge rank algorithms mould user conduct by imposing constant threats of invisibility and obsolescence, while Bivens (2017) has detailed how *Facebook's* software "normalizes a binary [gender] logic that regulates the social life of users" (894). Bivens argues that *Facebook's* 2014-custom gender project only at a surface level enables users to choose non-binary gender identities, while "the deeper level of the database" (894) builds binary gender profiles grounded in metadata such as the sex of repeatedly used personal pronouns.

Indeed, the often-assumed potentials of the digital to enhance self-expression, participation, and redistributions of power need to be critically reassessed. Fenton and Barassi (2011) for instance

argue that both Stiegler's (2008) notion of new media as enabling new forms of articulation that lead to an individuation of subjects through shared performances of identity and belonging (acting out), and Castells's (2009) ideas about the inherently beneficial role of increased participation in contemporary network societies, are based on a reductive understanding of identity, agency, and participation. Fenton and Barassi (2011: 191) write:

The problem with the notions of creative autonomy and individuation forwarded by Stiegler and Castells is that they prioritize individual agency over political and ideological context and resist problematizing the notion of autonomy therein. Autonomy in neo-liberal contexts may be guided principally by ego-centred needs and practices structured around the self that may implicitly endorse individualized and fragmented responses – a further push away from a collective public citizenry to isolated, atomized self-hood.

It seems that, besides supporting individual expression and self-presentation, digital networking technologies also serve the realisation and sedimentation of ideologically biased identity-potentials. Seen in this light, what appears as authentic expression of individual identities and selfhood on social media, might in reality rather resemble the design-driven constitution and reproduction of a particular version of an (a)political subject in line with neoliberal interests and practices (Fenton and Barassi, 2011; Thayne, 2012; Gehl, 2014; Lanchester, 2017). In the words of Gehl (2014: 23), '[s]ocial media is an instantiation – albeit a nascent one – of noopower: the action before the action that works to shape, modulate, and attenuate the attention and memory of subjects'.

Understanding the mixed impacts of digital technologies in general, and of social media in particular, requires a reconceptualization of identity. Rather than being tied to particular bodies,

the self becomes conceivable as an assemblage - a distributed, networked self that constantly emerges at various intersections between humans, non-humans, objects, materials, and energy flows (Coole and Frost, 2010; Papacharissi, 2011; Thayne, 2012; Banks, 2017). As Banks (2017) puts it, opposed to both Romanticist notions of an ultimately unknowable self and to a modern notion of a unified, objectively discernible self, the 'postmodern self' (421) is characterized by contingency and emerges in and through various multiplicities as 'a network of many different kinds of things that are linked across spaces' (423). In de-privileging the human and re-inserting it into complex socio-technical networks as just another object with certain agentic capacities, individual and collective identities become conceivable as fluid, hybrid, and constantly evolving - the always only partial and temporary results of continuous processes bringing together humans, objects, energy flows, and technologies (Coole and Frost, 2010).

In line with the thinking of for instance Fenton and Barassi (2011), Thayne (2012), Harcourt (2015), and Bivens (2017), it can be argued that the various technologically facilitated instantiations of a post-modern networked self in the sense of Banks (2017) to a large degree remain contingent upon neoliberal frames built into user interfaces of in particular commercial social networking applications. When seen from this vantage point, the user-led re-appropriations of *Foursquare's* implicit archival functionalities described by Saker and Evans (2016) emerge as always already pre-disposed by overarching neoliberal logics enshrined in the design of this technology.

The extent to which the interface design of social media has transformed users into transparent and malleable objects for automated surveillance and politically as well economically motivated interventions has lead Harcourt (2015) to abolish the much used Foucaultian panopticon as a diagram for what he terms 'our expository society in the digital age' (107). Instead, he proposes

the mirrored glass pavilion as a suitable alternative: ‘Partly crystal palace, part high-tech construction, partly aesthetic and partly efficient, these glass and steel constructs allow us to see ourselves and others through mirrored surfaces and virtual reflections’ (107). Rather than merely opening new venues for voice and self-representation, digital technologies also refract and reframe what is made to appear as mere reflection. Having thus ‘torn down the conventional boundaries between governing, commerce, and private life’ (187), the increasingly ubiquitous social networking technologies of the digital era profoundly challenge and change received notions of subjecthood, communality, and identity.

Social networking sites not only provide new means for participation, self-archiving, and articulation of reified individual identities, but also shape and mould contingent potentials for selfhood (Fenton and Barassi, 2011; Thayne, 2012; Gehl, 2014; Harcourt, 2015; Lanchester, 2017; Cheney-Lippold, 2017). This latter effect, however, is not only the result of an inherently political interaction design of user-driven surface archives that privileges quantification, instrumental relations and fragmentation, and that channels individual performances and expressions of identity into ultimately ideological pathways in real-time. In addition, corporate-owned and state-controlled deep archives capture and mine data implicitly produced in and through this real-time interaction, process it, and feed customized recommendations back to users giving rise to what Cheney-Lippold (2011) terms a *new algorithmic identity*, a type of ‘identity formation that works through mathematical algorithms to infer categories of identity on otherwise anonymous beings’ (165). The following section will take a closer look at this implicit deep dimension of digital archives.

User-extracted content: The state- and corporate-controlled deep dimension of iArchive

Having treated the interferences between user-controlled surface archives, digital design, and networked identities, the present part will show how both states and multinational corporations exploit user interaction in and with digital environments to create and curate their own deep archives that capitalize upon user-generated content and metadata. In this process, the networked self is dispersed across spread sheets, tables, and databases coalescing into a series of data-doubles that are formed in and through algorithm-driven predictive analytics, and that entail performative impacts on the lived lives of actual subjects.

Users interacting with contemporary networked environments operate in what Andrejevic (2007) has termed ‘digital enclosures’- virtual spaces ‘where every action and transaction generates information about itself’ (2) that is captured, mined, and instrumentalized. Augmented reality, geo-tracking, and location-based media increasingly mesh online and offline domains and bring ever-new categories of data into the purview of these processes. The constant mappings of activities across an increasingly comprehensive range of everyday practices produce a variety of user-profiles that each reflect the sources from which they were drawn. These data-doubles are intrinsically connected to actual subjects, but at the same time with necessity remain partial and contingent, merely pointing to certain potentials for identity. In a circular practice of algorithmic identity formation (Cheney-Lippold, 2011), the identified potentials are fed back to users in form of customized offers, suggestions, limitations, or other that operate upon the conduct of these users systematically inviting certain reproductive performances while demotivating others (Gehl, 2014; Pöttsch, 2015b; Harcourt, 2015).

Ernst (2013) notes that ‘through algorithms they [digital archives] are accessible to mathematical operations, something unprecedentedly new compared to the silence of the classical archive’ (86).

However, automatically assembled and harvested digital archives - and the implicit iArchives of social networking applications in particular - are not only accessible to regular mathematical operations. The sheer amount of data extracted from users makes these archives inaccessible to any other but algorithm-driven analysis. This reliance upon complex machinic operations and procedures has a variety of notable consequences.

As Gillespie (2014) has shown, algorithms are complex phenomena. In a very general definition they resemble ‘procedures for transforming input data into a desired output, based on specified calculations’ (167) and are, as such, not necessarily connected to computers. Computers, however, enable the application of very complex algorithms to data sets of a scale inaccessible to manual human operations. According to Gillespie, such complex, opaque, and often corporate-owned algorithms form a core operational frame for our engagement with digital networks and their implicit databases and archives. As such, they resemble a specific ‘knowledge logic’ with significant political, societal, and cultural ramifications. Gillespie introduces the term ‘public relevance algorithms’ (168) to highlight the increasing saliency of such complex mathematical operations.

Gillespie’s (2014) aim is to perceive public relevance algorithms as more than ‘abstract, technical achievements’ (169) by unpacking ‘the warm human and institutional choices that lie behind these cold mechanisms [algorithms]’ (169). By these means, he inserts a notion of contingency into the debate on algorithms and algorithmic power that can account for such factors as the selection processes behind the formation of data sets, the choices forming criteria of relevance, the practical appropriation of (often-assumed) algorithmic logics by active users, the discursive operations framing algorithms as objective and trustworthy, as well as potential constitutive impacts on consumer choices as well as on public and individual self-perception. In particular the latter two

elements, summarized by Gillespie under the headers of ‘cycles of anticipation’ (172-175) and ‘production of calculated publics’ (188-191) are of particular relevance to the present inquiry that aims to better understand the impacts of algorithm-driven analytics on processes and practices of identity formation in digital environments.

Also other scholars have taken a critical stance towards the role of algorithm-driven big data analytics in contemporary society and politics. Andrejevic (2013), for instance, argues that these practices privilege the identification of abstracted correlations and, therefore, de-emphasize context, contingency, and interpretation leading to a ‘post-comprehension era of information processing’ (35). Similarly, Boyd and Crawford (2012) have cautioned against hidden assumptions undermining the validity of big data-based methods, while O’Neil (2016), Foucault-Welles (2014), and Clough (2016) have warned that in-built tendencies of big data-based methods might discriminate against or make invisible non-normative identities and practices.

Raley (2013) draws upon the use of algorithms and big data for security measures to provide examples for what she terms performative impacts of these technologies. She writes that ‘the composition of flecks and bits of data into a profile of a terror suspect, the regrounding of abstract data in the targeting of an actual life, will have the effect of producing that life, that body, as a terror suspect’ (128). In a similar manner, Pöttsch (2015b) concludes his study on iBorder arguing that ‘the patterns [in sets of big data] giving rise to categories such as trusted traveller or terrorist threat are not necessarily revealing actual features of the world but form the basis for the practical implementation of measures that provide posthoc relevance to these patterns’ (114).

Moving from the field of terror and abstracted patterns-of-life analysis in drone warfare and border controls to the more mundane subject of day-to-day decisions and consumer choices on

social media, it becomes apparent that digital technologies shape and mould subjectivities at a variety of levels ranging from political convictions and interactions, via selections of friends, goods, and networks, to desire for affective commodities, or fear of ever-new potential threats. In all cases, highly customized offers and neatly tailored proposals that are attuned to specific profiles, combined with the constant ability to measure and track performances, movements, and responses, invite, and indeed create, particular desires, preferences, and behaviours in line with hegemonic interests. As Harcourt (2015: 217) sums up the situation with reference to social media, ‘for many of us, the new digital technologies have begun to shape our subjectivity’.

Given the contingent nature of a distributed self that constantly emerges in and through complex assemblages (Papacharissi, 2011; Cheney-Lippold, 2011, 2017; Banks, 2017), digital archives in general, and the iArchive of social networking sites in particular, become conceivable as sources of an algorithm-driven constitution of identities at both individual and collective levels. The data-doubles emerging from the various corporate-controlled big data repositories of social media companies not only secure advertising revenues, but also feed back into real lives and entail palpable material consequences. By means of such carefully attuned cybernetic feedback loops, these techniques frame reproductive performances and, in essence, gradually become constitutive of what they allegedly merely reflect.

What becomes apparent at this point, then, is that the primary function of the archive as a repository governing access to the past has indeed changed. Accompanying this past-bound archive is a new one that is directed at contingent futures – an implicit iArchive with the ambition to algorithmically presage and, indeed incite, probable or merely possible actions and performances that ultimately shape the world in its image. The term predictive retention serves to terminologically capture these affordances.

Predictive retention: Shaping the future by capturing the past

In her commentary *Datafied Citizens*, Barassi (2016) writes that ‘instead of focusing on the notion of surveillance alone, [...] we need to consider how these [digital] technologies enable a process of *profiling*, which enables the gathering of users’ past choices and behaviours to predict future needs’ (497; emphasis in original). The term predictive retention, refers to the techniques that enable such interconnections of assembled pasts with contingent futures. This technique, however, is not only about identifying future needs, but also entails a performative dimension that transforms knowledge about the past into reproductive social behaviour that takes part in actively forming the future in line with initial predictions.

Predictive retention employs complex algorithms to create user-profiles based on past behaviours and preferences logged at the implicit levels of digital archives. This knowledge of past patterns and tendencies enables future-bound interventions that use customized feedback loops to operate upon the conduct of subjects and, this way, shape and mould emergent subjectivities that then (re)produce the social world in correspondence with initial pokes and nudges. In Gillespie’s (2014: 174) terms, algorithm-driven ‘cycles of anticipation’ invite users to selectively ‘formalize’ themselves into ‘knowable categories’ adjusting their actual performances accordingly. Given the scope of user data within the purview of social media giants behind such applications as *Facebook*, *Twitter*, or *Instagram*, it becomes apparent that the corporate-owned deep dimension of iArchive constitutes a particularly valuable resource in this matter.

Indeed, the distributed and networked subject of a digital era (Banks, 2017), enmeshed in the ‘mirrored glass pavilion’ (Harcourt, 2015, 107) of social networking sites, is apparently both reflected and refracted in the various data-doubles emerging from the algorithmic profiling of

captured user-data. The co-constitutive impact of predictive retention, that both represents and shapes subjectivities, points to a form of power in the sense of Foucault (1982) that is not only coercive and limiting, but also productive of subjectivities and agencies, and that operates at the micro-level of everyday practices. Combining Foucault's (2004) notion of biopolitics with Deleuze's (1992) thinking on 'dividuals' and societies of control, Pöttsch (2015: 115) concludes his study on iBorder:

Power is no longer productive of docile individual bodies alone, but also of digitized data-doubles, or 'dividuals', whose contingent identity potentials entail performative sociopolitical effects that feed back into the bodies, subjectivities, and agencies they originated from.

Predictive retention through future-bound algorithmic analytics of users' captured pasts is a key operational dynamic of such processes and merits continued critical attention.

Conclusion

The present article has made a theoretical contribution and interrogated some of the shifting dynamics of contemporary digital culture and capitalism. Drawing upon classical notions of the archive and critical approaches to the political economy of the Internet, I directed attention to new technologically afforded practices and frames for the gathering, management, and curation of information in new media ecologies. Identifying a knowledge gap in important approaches to the digital archive, I developed the terms iArchive and predictive retention to enable a better understanding of the socio-political ramifications of data gathering, analysis, and commodification strategies of in particular commercial social media applications.

Homing these questions in on issues of identity and power, I argued that social media only in certain areas and only to a certain extent empower subjects and, in reality, cede enormous power

to private actors and states. This power is vested in exclusive rights to access and process data stored in the implicit deep dimension of social media memory (iArchive). The algorithm-driven analysis of logged user data gives rise to a variety of data-doubles that, through cybernetic feedback loops, offer nudges and pokes tailored to emergent profiles as such systematically operating upon the conduct of networked subjects and inviting performances that shape the social world in line with initial predictions. By these means, I argue, iArchive enables a form of predictive retention - a technologically facilitated future-bound practice of memory that becomes productive of what it allegedly merely reflects.

References

- Andrejevic M (2007) *iSpy: Surveillance and Power in the Interactive Era*. Lawrence: The University Press of Kansas.
- Andrejevic M (2013) *Infoglut: How too much Information is Changing the Way We Think and Know*. London: Routledge.
- Assange J (2015) Introduction: WikiLeaks and Empire. In: Assange J (ed) *The WikiLeaks Files: The World According to US Empire*. London: Verso, pp.1-19.
- Banks J (2017) Multimodal, Multiplex, Multispatial: A Network Model of the Self. *New Media & Society* 19(3): 419-438.
- Barassi V (2016) Datafied Citizens? Social Media Activism, Digital Traces and the Question about Political Profiling. *Communication and the Public* 1(4): 494-499.
- Bivens R (2017) The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook. *New Media & Society* 19(6): 880-898.
- Boyd D and Crawford K (2012) Critical Questions for Big Data. *Information, Communication & Society* 15(5): 662-679.

Pöttsch, *New Media & Society* (forthcoming), iArchive and Predictive Retention

Bucher T (2012) Want to be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook. *New Media & Society* 14(7): 1164-1180.

Carter RGS (2006) Of Things Said and Unsaid: Power, Archival Silences, and Power in Silence. *Archivaria* 61: 215-233.

Castells M (2009) *Communication Power*. Oxford: Oxford UP.

Cheney-Lippold J (2011) A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. *Theory, Culture & Society* 28(6): 164-181.

Cheney-Lippold J (2017) *We Are Data: Algorithms and the Making of Our Digital Selves*. New York: New York University Press.

Chun WHK (2006) *Control and Freedom: Power and Paranoia in the Age of Fibre Optics*. Cambridge: MIT Press.

Clough PT (2016) Rethinking Race, Calculation, Quantification, and Measure. *Cultural Studies* ↔ *Critical Methodologies* 16(5): 435-441.

Coole D and Frost S (2010) Introducing the New Materialisms. In: Coole D and Frost S (eds) *New Materialisms: Ontology, Agency, and Politics*. Durham: Duke University Press, pp.1-43.

Cook T (2013) Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms. *Archival Science* 13(2): 95-120.

Dafoe A (2015) On Technological Determinism: A Typology, Scope, Conditions, and a Mechanism. *Science, Technology & Human Values* 40(6): 1047-1076.

Deleuze G (1992) Postscript on the Society of Control. *October* 59: 3-7.

Derrida J (1995) *Archive Fever: A Freudian Impression*. Chicago: The University of Chicago Press.

Drucker J (2010) *Data as Capta*. Au: Druckwerk Rheintal GmbH.

Ernst W (2003). *Im Namen von Geschichte: Sammeln-Speichern-Erzählen*. München: Fink.

Pötzsch, *New Media & Society* (forthcoming), iArchive and Predictive Retention

Ernst W (2007). *Das Gesetz des Gedächtnisses: Speichermedien als Übertrag des 20. Jahrhunderts*. Berlin: Kulturverlag Kadmos.

Ernst W (2013) *Digital Memory and the Archive*. Minneapolis: University of Minnesota Press.

Fenton N and Barassi V (2011) Alternative Media and Social Networking Sites: The Politics of Individuation and Political Participation. *The Communication Review* 14: 179-196.

Foucault M (1982) The Subject and Power. In: Dreyfus HL and Rabinow P (eds) *Michel Foucault: Beyond Structuralism and Hermeneutics*. Brighton: The Harvester Press, pp.208-226.

Foucault M (2004) *Society Must Be Defended: Lectures at the Collège de France 1975-76*. London: Penguin Books.

Foucault-Welles B (2014) On Minorities and Outliers: The Case for Making Big Data Small. *Big Data & Society* 1(1): 1-2.

Fuchs C (2012) The Political Economy of Privacy on Facebook. *Television & New Media* 13(2): 139-155.

Fuchs C (2017) *Social Media: A Critical Introduction. Second Edition*. London: Sage.

Gehl RW (2014) *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism*. Philadelphia: Temple UP.

Gillespie T (2014) The Relevance of Algorithms. In: Gillespie T, Boczkowski PJ, Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge: MIT Press, pp. 167-193.

Greenwald G and Klein N (2016) Is Disclosure of Podesta's Emails a Step Too Far? A Conversation with Naomi Klein. *The Intercept* (October 19). <https://theintercept.com/2016/10/19/is-disclosure-of-podestas-emails-a-step-too-far-a-conversation-with-naomi-klein/>

Pöttsch, *New Media & Society* (forthcoming), iArchive and Predictive Retention

Greetham D (1999) 'Who's in and Who's out? The Cultural Poetics of Archival Exclusion. *Studies in the Literary Imagination* 32(1): 1-28.

Haebich A (2016) Fever in the Archive. *Thesis Eleven* 135(1): 82-98.

Handley T (2013) P2P searches as an Alternative to Google: Recapturing Network Value through Decentralized Search. *Journal of Peer Production* 3. <http://peerproduction.net/issues/issue-3-free-software-epistemics/peer-reviewed-papers/>

Harcourt BL (2015) *Exposed: Desire and Disobedience in the Digital Age*. Cambridge: Harvard UP.

Harrison S (2015) Indexing the Empire. In: Assange J (ed) *The WikiLeaks Files: The World According to US Empire*. London: Verso, pp.145-158

Herrman J (2016) Inside Facebook's (Totally Insane, Unintentionally Gigantic, Hyperpartisan) Political Media Machine. *New York Times Magazine* (August 24). http://www.nytimes.com/2016/08/28/magazine/inside-facebooks-totally-insane-unintentionally-gigantic-hyperpartisan-political-media-machine.html?_r=0

Hogan M (2015a) Facebook Storage Centres as the Archive's Underbelly. *Television & New Media* 16(1): 3-18.

Hogan M (2015b) The Archive as Dumpster. *Pivot* 4(1): 6-38.

Hogan M (2015c) Data Flows and Water Woes: The Utah Data Center. *Big Data & Society* 2(2): 1-12.

Joyce P (1999) The Politics of the Liberal Archive. *History of the Human Sciences* 12(2): 35-49.

Lanchester J (2017) You Are the Product. *London Review of Books* 39(16): 3-10.

Lynch M (1999) Archives in Formation: Privileged Spaces, Popular Archives and Paper Trails. *History of the Human Sciences* 12(2): 65-87.

Lyon D (2014) Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data & Society* 1(2): 1-13.

Pöttsch, *New Media & Society* (forthcoming), iArchive and Predictive Retention

Manoff M (2004) Theories of the Archive Across the Disciplines. *Portal: Libraries and the Academy* 4(1): 9-25.

Manovich L (2001) *The Language of New Media*. Cambridge: MIT Press.

O'Neil K (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin.

Osborne T (1999) The Ordinarity of the Archive. *History of the Human Sciences* 12(2): 51-64.

Papacharissi Z (2011) Conclusion: A Networked Self. In: Papacharissi Z (ed.) *A Networked Self: Identity, Community, and Culture on Social Network Sites*. London: Routledge, pp.304-318.

Parikka J (2013) Archival Media Theory: An Introduction to Wolfgang Ernst's Media Archaeology. In: Ernst W *Digital Memory and the Archive*. Minneapolis: University of Minnesota Press, pp.1-22.

Pöttsch H (2017) Media Matter. *TripleC* 15(1): 148-170.

Pöttsch H (2015a) The Emergence of iWar: Changing Practices and Perceptions of Military Engagement in a Digital Era. *New Media & Society* 17(1): 78-95.

Pöttsch H (2015b) The Emergence of iBorder: Bordering Bodies, Networks, and Machines. *EPD: Society & Space* 33(1): 101-118.

Qiu, JL (2016) *Goodbye iSlave: A Manifesto for Digital Abolition*. Urbana: University of Illinois Press.

Raley R (2013) Dataveillance and Counterveillance. In: Gitelman L (ed) *Raw Data is an Oxymoron*. Cambridge: MIT Press, pp.121-145.

Richards T (1993) *The Imperial Archive: Knowledge and Fantasy of Empire*. London: Verso.

Saker M (2017) Foursquare and Identity: Checking-In and Presenting the Self Through Location. *New Media & Society* 19(6): 934-949.

Pöttsch, *New Media & Society* (forthcoming), iArchive and Predictive Retention

Saker M and Evans L (2016) Locative Mobile Media and Time: Foursquare and Technological Memory. *First Monday* 21(2), n.p.

<http://firstmonday.org/ojs/index.php/fm/article/view/6006/5200>

Scahill J and Greenwald G (2016) Death by Metadata. In: Scahill J (ed) *The Assassination Complex: Inside the Government's Secret Drone Warfare Program*. London: Serpent's Tail, pp.94-107.

Stiegler B (2008) *Acting Out*. Stanford: Stanford UP.

Stoler AL (2009) *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton: Princeton UP.

Thayne M (2012) Friends Like Mine: The Production of Socialised Subjectivity in the Attention Economy. *Culture Machine* 13: 1-23.

Van Dijck J (2007) *Mediated Memories in the Digital Age*. Stanford: Stanford UP.

Van Dijck J (2014) Datafication, Dataism, and Dataveillance: Big Data Between Scientific Paradigm and Ideology. *Surveillance & Society* 12(2): 197-208.

Walker Rettberg J (2014) *Seeing Ourselves Through Technology*. Basingstoke: Palgrave Macmillan.

ⁱ *WikiLeaks* for instance has curated the Public Library of US Diplomacy (PlusD), a freely accessible, fully indexed and searchable online archive that is distributed across a variety of servers and peers (Harrison, 2015). It contains batches of secret diplomatic cables sent from US embassies around the world to the US State Department that were originally revealed to *WikiLeaks* by whistle-blower Chelsea Manning. Functioning as a globally accessible digital counter-archive on US foreign policy, PlusD enables 'an analysis of international relations that has not been hobbled by the censorship of classified materials' (Assange, 2015: 11). As such, it extends Lynch's (1999) technologically facilitated 'popular archive' (75) of a broadcast era into the realm of distributed digital networks pointing to a counter-hegemonic appropriation of these technologies. PlusD can be accessed here:

<https://search.wikileaks.org/plusd/about/>

ⁱⁱ For overviews see: [https://en.wikipedia.org/wiki/Global_surveillance_disclosures_\(2013%E2%80%93present\)](https://en.wikipedia.org/wiki/Global_surveillance_disclosures_(2013%E2%80%93present)) and <https://theintercept.com/documents/>.

ⁱⁱⁱ The algorithm-driven logics identified for the iArchive of social media applications retain their relevance for analysis of other digital archives such as those pertaining to for instance patterns in road traffic or the distribution of health problems. In all cases, the often corporate-controlled implicit levels of retention have ambiguous implications. As such, massively gathered traffic and health data is not only beneficial for road planning and the identification of potential health hazards, but can also be used to develop profiles with great fiscal value to for instance insurance companies that are put into the position to tailor products to the profiles of selected potential customers. Due to space limitations, these avowedly interesting, directions cannot be pursued any further in the present article.