The Datafied Society

Published by Amsterdam University Press

The Datafied Society: Studying Culture through Data.
Amsterdam University Press, 2017.
Project MUSE. muse.jhu.edu/book/66554.

For additional information about this book
https://muse.jhu.edu/book/66554
Humans have long defined, assessed, analysed and calculated data as factors in how they navigate reality. Indeed, the rules for what constitute data, together with the logics of their assembly, make up a core component of culture. Whether they be omens or numbers, whether they are qualitative or quantitative, whether they involve heuristics, hermeneutics or the rules of mathematics, the dyad of data and their organizing schemes give cultural eras their specificity. Considering developments ranging from Mayan astronomical calendars to Copernicus’s heliocentric observations, from seventeenth-century navigational charts to twentieth-century actuarial tables, one might say that this dyad underpins cultural possibility itself.

Data have never been more abundant than they are today. Their unprecedented quantity owes as much to the digital encoding of most traceable phenomena as to the production of data by actors beyond our species. Whereas in the past, human observation translated events in the world into data, today, networked non-human actors are capable of directly generating machine-readable data. But as in the past, all that data would be meaningless without an organizing scheme. Behind the quintillions of bytes, behind our computers’ ever-growing processing power, is an organizing scheme in the form of the algorithm. Like data, algorithms can be human- or machine-generated. And although an ancient idea, the algorithm has – or so I will argue – reached a tipping point in terms of its cultural operations: it is now being deployed in ways that redefine long-held subject-object relationships and, in so doing, it poses some rather fundamental epistemological questions.

This change in the balance of things has produced its share of anxieties, as familiar ways of doing things seem superseded by ‘the algorithm’. The recent explosion of headlines where the term ‘algorithm’ figures prominently and often apocalyptically suggests that we are re-enacting a familiar ritual in which ‘new’ technologies appear in the regalia of disruption. But the emerging algorithmic regime is more than ‘just another’ temporarily unruly
new technology. My thesis is that the algorithm, an approach to problem solving that goes back at least to Euclid’s *Elements* (ca. 300 BC) and that enjoyed significant development in the hands of Leibniz and Pascal, has achieved new force as a cultural technology thanks to a confluence of factors that include the emergence of big data, intensive processing power and high-speed networks. It alters the subject-object relationship hard-wired in the project of the modern and visible in technologies like the printing press and three-point perspective, both of which amplify individual agency. Yet, like these technologies, the algorithm, combined with data, can be read as defining an emerging epistemic era. If we are indeed like those in the early fifteenth century who were poised on the edge of a new order of things, will we, like some of them, be inclined to embrace their potential for a new vision of ourselves in the world, a new social order? Or will we miss the radical potential of a new technology, retrofitting it to serve the still-dominant interests of the old?

Technologies do not, in themselves, change anything, but rather are socially constructed and deployed. So as we watch the possibilities of a new technology take shape in the hands of those with the greatest economic power, we have good reason to be anxious. The dyad of big data and algorithms can enable new cultural and social forms, or they can be made to reinforce the most egregious aspects of our present social order. That is a political choice, of course. But what is truly new about this configuration is that we have a choice at all, of a magnitude not seen since the fifteenth century. The pages ahead will chart these new enablements: first, by considering the definitional dynamics of algorithms; second, by looking at their newly acquired place particularly as a condition of cultural production; and finally, by raising some questions regarding the larger epistemic implications of this new order and how we as a society will grasp it.

**Definitional Dynamics**

The term ‘algorithm’ seems to conjure up responses disproportionate to the simplicity of its meaning. Formally speaking, an algorithm is simply a recipe, a process or set of rules usually expressed in algebraic notation. The actual values plugged into the algorithm are less the point than the step-by-step formulations that govern their processing. They scale easily, whether working with the relatively meager data of the pre-computer era or the more than 2.5 quintillion bytes of data generated daily at the time of this writing. Yet despite their relative simplicity, algorithms today
pose some significant definitional problems, mostly because of a series of misapprehensions. Tarleton Gillespie (2014) has noted three broad uses of the term that obscure its meaning. Algorithms are invoked as synecdoche when the term stands in for a sociotechnical assemblage that includes the algorithm, model, data set, application and so on. They reveal a commitment to procedure, formalizing social facts into measurable data and clarifying problems into models for solution. And they function as talismans when the term implies an ‘objectifying’ scientific claim. Indeed, one might step back and note that these three uses say much more about social anxieties and aspirations than they do about algorithms. How, for example, can one make a claim to ‘objectivity’ with an authored protocol whose operations depend on the highly variable character and structure of a particular data set? And yet a glance at any newspaper will confirm the accuracy of Gillespie’s insights about the term’s ambiguity. The definition of the algorithm is also complicated by more insistent epistemological problems. Nick Seaver (2013) finds that most discussions of algorithms get caught up with issues of access and expertise. Access is an issue because many commercial algorithms, Google’s for instance, are closely guarded secrets. ‘If only we had access…’ the mantra goes. But even if we had access, we would immediately face the expertise problem, for most individual algorithms inhabit vast interdependent algorithmic systems (not to mention models, goals, data profiles, testing protocols, etc.), and disaggregating and making sense of them typically require large teams of experts. But even more troublesome is the fact that any given process usually has many possible algorithmic combinations (circa 10 million in the case of a Bing search), some of which might be uniquely deployed or used for purposes of personalization or even testing. Individual algorithms and algorithmic clusters are recycled and appear in different settings, with some dating from before World War II still in circulation today. This means that we can never be sure precisely which set of algorithmic elements we are examining, and even if we were, the work of personalization would limit our ability to compare findings. A further twist appears in the form of disciplinary specificity. The valences of the term ‘algorithm’ differ in mathematics, computer science, governance, predictive analytics, law and in the culture at large, complicating cross-disciplinary discussion.

Finally, unlike earlier technologies, developments in machine learning have enabled algorithms to self-optimize and generate their own improvements. They can now self-author and self-create. This greatly complicates notions of authorship, agency and even algorithms’ status as tools, which imply an end user.
Together, the various factors described by Gillespie and Seaver, which are embedded in our tradition of study and even our inherited notions of agency and authorship, all combine to render the simple definition of an algorithm as a ‘rule set’ or ‘recipe’ into something quite… loaded. And they fundamentally challenge our inherited notions of culture and cultural production. The humanities research agenda not only has to deal with the implications of radically reconfigured notions of the author, agency and textual stability, but also has to embrace radically expanded corpora. Data, the structure of the data set, models, software systems and interfaces all play determining roles in cultural production and, as such, are not only appropriate but increasingly important sites for humanistic inquiry. Their analysis requires not only new literacies but evaluative paradigms that in many cases have yet to be created. Lev Manovich (2001) made an early appeal to meet these needs in his *The Language of New Media*, and this essay extends that call to include the algorithms underlying these operational systems.

**Culture**

Given the role that the dyad of algorithms and data currently plays in shaping our access to information (Google) and the social world (Facebook), and their centrality to finance (algorithmic trading) and governance (from predictive policing to NSA-style parsing of vast troves of data), looking at their cultural work might seem a low priority. Each of these sectors reveals some affordances of the pairing, and their most visible – and disturbing – applications reflect the interests of the prevailing power structure. However, the abusive deployment of algorithmically enabled data says more about the contradictions of our social order than the algorithm or data per se. Blaming ‘the algorithm’ or ‘big data’ puts us in the position of a bull fixated on the matador’s cape: we fail to see the real source of malice.

We can sidestep the easy conflation of algorithms and data in the explicit service of power by turning to the cultural sector in order to throw into relief the dyad’s capacities to re-order the subject-object relationships at the heart of the new representational order. This re-ordering has far more profound implications than the retrofitting of algorithms and data in the service of twentieth-century notions of power (though doing the latter may wind up killing us if we aren’t alert). With art it is generally easier to see through the representation process and find traces of the underlying production system. The arts help us to see more clearly.
Just as algorithms have a deep history but have also recently achieved new power thanks to their changing circumstances (big data and dramatic improvements in processing and transmission), their use in the arts also has a long history as well as a fast-evolving present. The historically oldest applications deploy algorithms to organize data for purposes of textual production, basically using the algorithm like a traditional artistic tool, though with an important twist. More recent applications go further, using algorithmic configurations of data for purposes of textual selection and customization, combing through large data sets to establish correlations regarding taste and likely matches between users and texts. Brief disambiguations of these different applications follow:

**Algorithms as Tools in Traditional Artistic Production**

The canon form in music, essentially an algorithm, goes back at least to the Middle Ages; and algorithms have appeared in works ranging from the *Musikalisches Würfelspiel* attributed to Mozart to Lejaren Hiller’s compositions using the ILLIAC computer in the 1950s. Brian Eno (1975) put his finger on the aesthetic twist of this application when he said:

> Since I have always preferred making plans to executing them, I have gravitated towards situations and systems that, once set into operation, could create music with little or no intervention on my part. That is to say, I tend towards the roles of planner and programmer, and then become an audience to the results. (n.p.)

This disaggregation of artistic process from execution is nothing new (Rodin famously relied on it for his major sculptural works) but it has served as a persistent characteristic in the long history of algorithmic art. Tradition enters the picture when artists make a claim for ‘their’ authorship, rendering the algorithm a tool. The 1968 exhibition Cybernetic Serendipity, with its display of algorithmically generated music, painting, choreography, film and graphics, demonstrated the powers of this new toolkit across the arts to audiences in London and Washington. By the mid-1990s, artists such as Roman Verostko and Jean-Pierre Hébert proclaimed the tool as the basis of a movement: the Algorists.² Today, the integration of algorithms into everyday textual production is so fundamental as to be quotidian (algorithms enable colour

---

² See www.algorists.org/.
correction, editing and the very existence of image in film and video (Hoelzl & Marie 2015); recording, mixing and the creation of sound elements in music; the word processing program that I am using to craft this text, and so forth).

But paired with big data, the algorithm has grown far more powerful. According to The New York Times, the company Automated Insights alone created more than one billion algorithmically generated stories in 2014, mostly routine sports and financial market reporting (Podolny 2015). These two domains are made up of well-structured data sets, with timelines and data points that enable easy characterization and serve as low-hanging fruit for an emergent industry. But The Times story gave a sense of the ambitions for storytelling algorithms produced by companies such as Narrative Science, and they go far beyond ‘translating’ simple data trails to generating creative prose and poetry.

Textual Recommendation Systems

A very different and relatively recent cultural use of algorithms paired with large data sets takes the form of selecting and pushing which texts we have access to, that is, of recommendation systems. Consider Echo-Nest’s prediction algorithms that comb through data derived from millions of users’ behaviours as well as data drawn from musical texts, seeking correlations by extrapolating from past behaviours into future desires or by searching for other users’ patterns that might offer a basis for suggestions. To the extent that users play along and offer consistent feedback, Pandora, Spotify and other streaming music services that use EchoNest’s algorithms demonstrate an uncanny ability to identify and provide access to the desired, the familiar and the reassuring. As users of Amazon’s book recommendation services or Netflix’s film and video suggestions know, the same principles apply on these platforms as well. Indeed, one of the often referenced developments in this space was the 2009 Netflix Prize, a $1,000,000 bounty for creating the greatest improvements to Netflix’s own collaborative filtering algorithm for predicting user ratings of films (the winner, Bellkor, achieved a 10.09% improvement on predictions). In these

---

3 The Netflix competition began in 2006 and ended with the 2009 award to the Bellkor team. The terms of the prize required that winners publish a description of the winning algorithm. Throughout the multi-year process, critics claimed that Netflix’s release of data sets violated US Fair Trade laws as well as the Video Privacy Protection Act. The Netflix Prize website has archived much of the process: www.netflixprize.com/.
predictive systems, the past is prologue, as the data generated through our earlier interactions shape the textual world selected for us. No ‘surprises’ or ‘unwanted’ encounters, just uncannily familiar themes and variations. This logic extends into the informational domain as well, where it has been the subject of sharper critique, mostly focused on the argument that such predictive systems create an echo chamber in which our existing views of the world are reinforced but rarely challenged.4

**Prediction as a Gatekeeper for Textual Production**

Taste prediction has another rapidly growing dimension: in some settings the combination of data and algorithms serves as gatekeepers for cultural production, and in the process has displaced the embodied knowledge of established tastemakers. Epagogix, a company that specializes in risk mitigation, has found a niche in advising investors in the film and television industry about the likely success of a given project. Data from the script as well as various casting configurations are analysed by Epagogix's proprietary algorithms, along with a financial assessment that may (or may not) serve as an incentive for investment. Needless to say, long-time industry specialists view such developments with suspicion if not outright contempt, but investors, convinced by the seeming objectivity of numbers and the system’s more-often-than-not accurate predictions, think otherwise. Such investor response is understandable at a moment when most stock trading is algorithmically determined: the algorithm is a vernacular of sorts. But it also confirms Gillespie’s observation that the algorithm is a talisman, radiating an aura of computer-confirmed objectivity, even though the programming parameters and data construction reveal deeply human prejudices. The bottom line here is that decisions regarding what will and will not be produced are now often based on data of unknown

---

4 The ‘echo chamber’ effect is widely used in journalism and mass communications to refer to the closed circle of media utterances and audience beliefs, as in Kathleen Hall Jamieson & Joseph N. Cappella, *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment* (New York: Oxford University Press, 2008). The concept has been extended to social media such as Facebook, where algorithms filter and sequence the posts that users see, effectively creating an echo chamber. See for example Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (London: Penguin Press, 2011). Facebook researchers Eytan Bakshy, Solomon Messing and Lada Adamic argue, however, that the data do not support this view (‘Exposure to Diverse Information on Facebook’ 7 May 2015: https://research.facebook.com/blog/exposure-to-diverse-information-on-facebook/).
quality (What do they actually represent? How were they gathered?) that are fed into algorithms modelled in unknown ways (What constitutes success?).

‘Live’ Textual Production

Another relatively recent application of algorithms and textual data sets regards what might be termed ‘live’ or dynamic ‘on-demand’ textual production. Whereas we saw that one of the oldest continuing cultural uses of algorithms was as a tool to streamline what we might term the production of traditional texts (from an occasional Mozart composition to Narrative Science’s articles for Forbes), here the texts are dynamic (in the sense of being interactive), responsive (in the sense of being tailored to individual preferences) and inherently unstable (that is, no two texts are identical). Interactive documentaries, often in the form of textual environments (i.e. databases) permit the user to follow his or her interests, with the resulting navigational trail as text. This approach requires the user’s active interaction and choice; however, we are fast moving towards a situation where choices regarding text selection (i.e. data selection) will be made on the fly by algorithms armed with data about our preferences. Here, personalization algorithms meet textual production algorithms to create what seems like a seamless, traditional text, even though it will be a unique, real-time data ensemble for our eyes only.

The hundreds of reader responses to the Times article amply demonstrated the provocative nature of these developments: text-generating algorithms force us to ask what it means to be human and how that relates to artistic production; production filters force us to reflect on the nature of our automated cultural gatekeepers; personalized texts force us to consider the future of shared experiences. For most commenters, the answer was clear-cut: algorithmic creativity and content-as-data in the traditional cultural sectors seem oxymoronic. Culture is precisely about human expression, and anything else is either trickery or parody. But to designers of algorithms and data sets, such discourse – to the extent that it articulates a human je ne sais quoi – is useful in pinning down precisely the gap between human and algorithmic expressions, enabling engineers to define and to chip away at the problem. Much like the issue of intelligence, long-held assumptions regarding man-the-measure are undergoing a Copernican-like decentring, and in this sense, the coincidental appearance of developments such as post-humanism, actor network theory, object-oriented ontology and the rest
suggests that sectors of the academy are indeed thinking seriously about a paradigm shift and alternatives to a human-centric culture. All this is to say that the cultural deployments of algorithms and data have different valances. An early and continuing strand of creativity has harnessed algorithms and data to the work of familiar artistic paradigms, where things like authorship and attribution are still relevant (Eno and Verostko still sign their computer-generated works). But as just noted, a rapidly emerging set of developments has seen algorithms used as filters, shaping our access to the cultural repertoire; as gatekeepers, helping to determine what will and will not be produced; and as semi-autonomous forces of production, writing texts, composing music and constructing films – all dynamically personalized and assembled on the fly. Of course, these are still early days and results can sometimes be erratic (Microsoft’s Tay AI neural net chatbot was abruptly terminated shortly after her public release in 2016 when ‘she’ spouted Nazi rhetoric (Bright 2016)). But generally, so long as Moore’s Law holds, these developments are growing more intensive, driven by the ever more pervasive place of computational systems in our lives, the ability of algorithms to self-improve without active human intervention, and the ever-increasing depth of our data sets. They raise crucial questions regarding agency and attribution (how to negotiate the space between human designers and machine learning? What is the nature of authorship and the creative act?), point of view (whose values, experiences and perceptions are bound up in this new order and the underlying definition of data?) and cultural access (what notion of ‘personalization’ enables – or delimits – our encounters with texts, and with what implications?).

The Bigger Picture

Why do these questions, and the increasing insistence with which they are posed, matter? What are the stakes involved? Heidegger (1938) used an image, the ‘world picture’ (Weltbild), to mark the birth of the modern, saying that the moment at which the world becomes picture is the same moment that the human emerges as the subject in a characteristically...
modern subject-object relationship. He argues that the modern social order can be defined through a representational system characterized by precisely defined subject-object relations (the world as picture), a metaphysics of exactitude and an underlying spatiotemporal grid – all qualities that we can see materialized in Gutenberg’s printing press and Brunelleschi’s notion of perspective, technologies that amplified the subject and her viewing position.\(^6\)

In the hundreds of years between these early-fifteenth-century developments and Heidegger’s twentieth century, despite countless historical undulations and discoveries, we encounter a consistent logic of attribution, of a stable self and relationship to the world, a notion of mathematics as a language of precision, calculability and predictability. By contrast, the algorithm enabled by big data – as shown in the cultural deployments just discussed – stands between the calculating subject and the object calculated; it refracts the subject-centred world. Together algorithms and data filter what we have access to, produce our texts with unseen hands and unknown logics, and reshape our texts, rendering them contingent, mutable and ‘personalized’.

The implications of this change, if we take thinkers like Heidegger seriously, are profound. Consider the contrast between Diderot’s *Encyclopédie* and the crowdsourced Wikipedia, or between Canaletto’s painting of Piazza San Marco and the hundreds of differently authored photos that in the aggregate constitute Photosynth’s ‘synth’ of the same locale. With Diderot’s compendium and Canaletto, the editor and the painter are known, their point of view embodied, their relationship to the object clear, and their text stable. With texts such as Photosynth, the author is potentially collective, diffused and anonymous; the points of view multiple; the relationship to the object both data-based and algorithmically mediated; and the text ever-changing and mutable. These differences, *grosso modo*, distinguish the project of the modern, Heidegger’s ‘age of the world picture’, from the enablements of the data-powered algorithmic era.

Authorship, in the algorithmic context, is both pluriform and problematic. It turns on the algorithmic re-ordering of data (textual elements), informed and shaped by algorithmic assessments of data (reliability and preference correlations), all algorithmically calculated to achieve certain data markers (user rates). This is not to ignore human agency: humans

---

6 For a developed version of this argument, particularly as it regards visualization technologies such as augmented reality and Photosynth, see William Uricchio, ‘The Algorithmic Turn: Photosynth, Augmented Reality and the State of the Image’ in *Visual Studies* 26:1 (2011): 25–35.
take the photographs or shoot the video clips that make up a Photosynth ‘synth’ or an interactive documentary (even if these processes also reveal fundamental partnerships between the human creator and light-as-data and algorithms-as-image-stabilizers built into our cameras). But when those human-created images are re-rendered into the abstraction of data sets, when those data are algorithmically deployed and stitched back into an image, then categories such as authorship, agency and motive are fundamentally blurred. Descartes’ triumphant subject and the Ich implied in Heidegger’s Weltbild are not eradicated, for their traces remain abundant in the individual images and clips. Rather, they are fundamentally repositioned by the algorithmic and data regimes that now stand between the subject and object.

If we understand this, we can think through the opportunities that await us rather than panic at the loss of the old certainties. We can explore the affordances of algorithmically-enabled collaboration and the new forms of collective creativity that might emerge from a world re-articulated as data, rather than tolerating the crude use of algorithmic systems and data sets to exploit and oppress. We can try to understand the implications of widespread personalization, the challenges of a predictive economy in which data trails become constitutive and the impact of a culture of radical contingency. How? We can first comprehend that conditions have changed, that we need to shift our focus from the simple causalities of the subject-object binary to a far less decipherable algorithmic intermediary. To do so requires a new literacy, not in the sense of making composited algorithms legible, for that is beyond comprehension, but rather by attending to their operations, noting their defaults, critiquing their judgements and the definitions of the data they are processing. And particularly at this moment of transition, we must carefully assess the ends to which these new tools are put – whether they are being bent to the whims of the old subject (aggregating power and control) or facilitate new collectivities.

In framing these issues, I’ve gone back several times to the fifteenth century and the emergence of modern technologies such as the printing press and three-point perspective. These technologies amplify the position of the individual human subject and resonate through the six centuries of the modern that followed. That they are taken for granted even today, when they are increasingly displaced by a radically different representational regime and set of technologies, shows us that old habits die hard. Not

---

7 And conversely, algorithms and data sets are in the majority of cases human-authored and assembled, even if they can go on to self-generate, further complicating the situation.
surprisingly for a moment of transition, today’s dominant deployments of algorithmically enabled systems emulate the representational traditions of the past: they often look ordinary and familiar. And we respond accordingly, reading a Narrative Science-authored newspaper article as we would a human-authored story or viewing a Photosynth image as we would a photograph. It’s easy to miss the radical reworking of cultural logics in media res, easy to re-inscribe the new and uncertain into the familiar categories of the past... or to reject them as threats to the status quo.

We can probably learn something from our predecessors in the late Middle Ages, poised on the cusp of the modern, encountering the printing press and three-point perspective. What did people make of new and, in retrospect, era-defining technologies before that era was defined? Scholars such as Elizabeth Eisenstein (1979) have tended to see the printing press as a trigger for the modern (knowledge stabilization, spread, etc.), while others such as Adrian Johns (1988) have more recently chronicled the disparate and unruly practices that attended its initial decades. In Eisenstein’s story, emerging technologies exerted an impact, making a splash as harbingers of the new; and in Johns’s, they were taken up by a late-medieval populace and used in aberrant and contradictory ways. I would argue that we are in a similar position with our new era-defining technologies of algorithms enabled by big data and massive processing power.

The new era has yet to be defined, and it is impossible to know how future historians will inscribe our trajectory. Of course, the ‘newness’ of this regime comes with the danger that it will be retrofitted to sustain the excesses and contradictions of the fast-aging modern, to empower particular individual points of view, to control and stabilize a master narrative. But it also offers an opportunity for critical thinking and an imaginative embrace of the era’s new affordances. And for these opportunities to be realized, we need to develop critical perspectives, to develop analytical categories relevant to these developments and our place in them.

Much of what we today call the humanities harkens back to traditions developed during the long span of the modern, traditions predicated upon the stable subject-object relationship noted earlier and captured by Heidegger’s concept of the Weltbild. But although the term ‘humanities’ was coined in the Renaissance of the fifteenth century (studia humanitatis), its texts and values go back to the pre-modern world of classical Greece and Rome, where the humanities involved practice more than study. The question is whether we can draw on this era-spanning tradition to anchor, critically assess and navigate the ‘age of the algorithm’ (to put the new era in terms equivalent to Heidegger’s ‘age of the world picture’). Can we disentangle the
centrality of the individual subject from the humanity at the core of the humanities agenda? Can we rethink our inherited categories of authorship and agency in ways that stimulate a critical discourse of collaborative and algorithmically-enabled work? Can we shift from familiar conditions such as precision, calculability and predictability and learn to grapple with the contingent, mutable and personalized? Is a poetics of data within our reach or even desirable?

To even begin to answer these questions, we need to develop new literacies capable of assessing various data forms and organizing schemes such as algorithms. We have to understand how they are deployed and develop a critical sense of their limits, capacities, implications and possibilities. How do they operate, not so much as technological ensembles but as patterning activities, as enablers of collaboration and creativity, as potentially critical practices? The humanities – the questions their texts and values pose, the comparative and historical framings they deploy, the analytic attention that they expend – have never been more important. Yes, modernist assumptions need fundamental revision, and their corpora need to be radically expanded to include categories like data and algorithms. And yes, poised as we are on the cusp of a new era, we have much to learn from similar transitions in the past, particularly regarding the predictable rear-guard actions of those who seek to exploit the potentials of the new to extend the power dynamics of the old.

References


