<http://www.diva-portal.org>

This is the published version of a paper published in *First Monday*.

Citation for the original published paper (version of record):

Eriksson, M. (2016)

Close Reading Big Data: The Echo Nest and the Production of (Rotten) Music Metadata

First Monday, 21(7)

<https://doi.org/10.5210/fm.v21i7.6303>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-148443>

Close reading big data: The Echo Nest and the production of (rotten) music metadata by Maria Eriksson

Abstract

Digital music distribution is increasingly powered by automated mechanisms that continuously capture, sort and analyze large amounts of Web-based data. This paper traces the historical development of music metadata management and its ties to the growing of the field of ‘big data’ knowledge production. In particular, it explores the data catching mechanisms enabled by the Spotify-owned company The Echo Nest, and provides a close reading of parts of the company’s collection and analysis of data regarding musicians. Doing so, it reveals evidence of the ways in which trivial, random, and unintentional data enters into the data streams that power today’s digital music distribution. The presence of such curious data needs to be understood as a central part of contemporary algorithmic knowledge production, and calls for a need to re-conceptualize both (digital) musical artifacts and (digital) musical expertise.

Introduction

In April 2015, the company Echo Nest revealed the staggering figure of having collected over 1.2 *trillion* data entries about music and artists on their Web page. The Echo Nest is an actor in the business of producing metadata, or what is commonly described as ‘data about data.’ In essence, this means that the company collects, synthesizes, and examines billions of data points regarding artists and music on a daily basis. The outcomes of this accumulation and analysis of data — what The Echo Nest itself calls “music intelligence” — later forms the basis of recommendation algorithms, automatic playlist generators, and personalized radio stations on the Web. It is such metadata that helps to power today’s digital music streams, and aids in transforming messy currents of digital music into what is often framed as ‘personalized’ and ‘trend-sensitive’ floods of content delivery. Placed at the center of the contemporary ‘big data deluge’ (Bowker, 2013). The Echo Nest is helping to propel the data-driven turn within the music industries, while creating new kinds of measurements and conceptualizations of music and artistry along the way.

Recently, much attention has been directed toward the larger paradigm shift where 'big data' is coming to underlie more and more of public and private life (Manovich, 2011; boyd and Crawford, 2012, 2011; Berry, 2011; Gitelman, 2013). Many have also shown interest in the functions of the algorithms that serve to make sense out of such amounts of data, and problematized how they are governed (Gillespie, 2012; Mager, 2012; Cheney-Lippold, 2011).

However, relatively little research has provided detailed accounts of what kind of materials that are turned into 'big data'. Likewise, few studies have looked closer at how such information is organized and made 'algorithm ready' (Gillespie, 2014). This article aims to contribute to such an understanding, and uses the example of music metadata management as a lens through which to explore how information is assembled and turned into metadata by algorithms and information managers.

The paper answers calls for qualitative approaches to big data (Burgess and Bruns, 2012; Crawford, 2013), and joins social anthropological, media scholarly, and media archeological discussions on digitized forms of knowledge production. It attempts to challenge big data's urge to be read in an aggregated form and instead take up on its intricacies. Thus, it looks at big data more through a magnifying glass, than through the zoomed-out perspective by which it is commonly portrayed. An experiment that lasted for one month and provided a snapshot of the growth, decline, and transformation of The Echo Nest's collection and evaluations of music metadata will function as the gateway into such an analysis.

In the process of close reading and examining this data I am more interested in the internal dimensions of algorithmic knowledge production, and less on its possible effects on music industry practices. Therefore, I ask: How does a company like The Echo Nest filter, augment, present, and make content meaningful for its addressees? In what ways does this kind of data management produce different kinds of 'knowing' about the world? And what can methods for selecting and aggregating data about music tell us about wider renegotiations of expertise under the current paradigm of big data?

To answer these questions, this article first roots large-scale systems for music metadata collection and analysis in histories and theories of metadata, discussions about the growing trust and faith in 'big data', and conversations on the increased reliance on software for the curation of information. In such a context, I argue that metadata is doing much more than serving as an informative background to music; it is transforming how digital objects should be conceptualized, since it is making boundaries between original sources and their data obsolete.

Against this backdrop, I later highlight some of the (more or less intended) outcomes of large-scale systems of metadata management, and take a look at how automatic (or semi-automatic), systems of information management sometimes produce curious forms of knowledge. Based on a longitudinal study of The Echo Nest's collection of metadata, I discuss some of the odd, surprising, and peculiar data found within their archive. Here, the concepts of "rotten" data (Lévi-Strauss, 2013; Boellstorff, 2014), and computational 'randomness' (Parisi, 2013) are used to think through the logics and outcomes of algorithmic expertise. By doing this, I hope to lift broader concerns about the types of knowledge produced by big data analysis.

Metadata, ‘big data’ & coded systems of curation

As mentioned before, metadata is commonly described as “data about data” — a kind of material that *surrounds* original sources, and in one way or another provide contextual reports about their existence in the world. While the term ‘metadata’ was introduced by computer scientists as late as in the 1980s, metadata itself is not something new (Caplan, 2003). For example, Hui (2014) reports that one of the first large-scale systems for metadata production and analysis — an arrangement of clay tablets revealing information about livestock — has been linked to the Sumerian civilization, active in the early Bronze Age [1]. Similarly, card catalogues (Krajewski, 2011), and other information systems used in libraries, can be seen as predecessors to the large-scale digitized metadata archives of today, as they too were entities that collected and stored ‘information about information’ in comprehensive and organized ways.

At the heart of the uses of metadata lies a desire to distinguish, categorize, and establish relationships between (and thereby also control) different kinds of people, objects, substances, and information. Metadata is commonly assumed to create pathways of discovery in massive amounts of data by providing an efficient system for the search and retrieval of content. Following the logic “the more the better” (Gitelman and Jackson, 2013; Snickars, 2014), it has been approached as a kind of information type that improves chances of unearthing new knowledge in large piles of information and increases interoperability; that is, the possibility to compare, interweave, and establish relationships between people and things.

While the management of metadata has traditionally been governed by information experts (archivists, museologists, secretaries, etc.), digitization has increasingly meant that such tasks are outsourced to software systems that automatically crawl, organize, label, group, analyze, and manage various kinds of digital content. Digitization has also meant that metadata management has taken the step into the sphere of ‘big data,’ where social action is turned into digital quantified data, and is consequently used to analyze, track and predict the behaviors and mindsets of people (Mayer-Schönberger and Cukier, 2013).

Big data has been described as a “cultural, technological, and scholarly phenomenon” that rests on the interplay of technology, analysis, and mythology [2]. Adding to such a description, it could further be stressed that the notion of ‘big data’ is an industrial phenomenon; and one which is infused with corporate overtone and interests (Puschmann and Burgess, 2014). With billions of data entries in its archive, The Echo Nest’s collection of metadata represents one of the big data archives that help to shape (and profit from) an emerging ‘dataverse’ in which more and more of our surroundings are powered by algorithm managers, and the kinds of ‘knowing’ extracted from large datasets (Bowker, 2013).

One discussion concerning big data that has recently received a lot of attention, is its purposed “rawness,” or conversely “cooked” nature (see Gitelman, 2013, including Bowker, 2013; Gitelman and Jackson, 2013). This conversation — which in many ways has functioned as a critique of the objectivity and legitimacy endowed to large and supposedly “raw” datasets — has largely echoed Claude Lévi-Strauss argument that in the sphere of cooking, there is no condition of pure rawness. Instead, Lévi-Strauss [3] has pointed out that all ‘raw’ things intended for digestion, also need to be “selected, washed, pared or cut, or even seasoned.” Following this logic, every instance of ‘producing’ data (both in the sense of analyzing or

‘cooking’ it, and labeling it as ‘raw’) is also a political act, embedded in culture and dynamics of power. Data generation is thereby inevitably seen as the outcome of the act of interpretation and manufacture; it is something which ascends from the choice of what to include and what to exclude, and how to value the included (Gitelman and Jackson, 2013; Bowker, 2013).

In making a contribution to this discussion, Boellstorff (2014) has referred back not to Lévi-Strauss original book *The raw and the cooked* (1969), but to *The culinary triangle* (2013) which Lévi-Strauss originally published one year later. In this text, Lévi-Strauss places the categories “raw” and “cooked” in a triadic relationship with the “rotten,” and in such a mode of theorization, Boellstorff (2014) explains that:

“‘raw’ and ‘cooked’ are not set in a binary where raw = nature and cooked = culture. Instead, they are framed as elements of a “culinary triangle” shaped by the intersection of the binarisms of nature/culture and normal/transformed.”

The introduction of the category “rotten” allows for considering the unintentional, spontaneous, uncontrolled, and random processes taking place within the making of big data, argues Boellstorff. It also avoids “imputing a timeline” to the way that data moves between stages such as ‘untouched’ and ‘managed.’ Instead, the term rotten “reflects how data can be transformed in parahuman, complexly material, and temporally emergent ways that do not always follow a preordained, algorithmic ‘recipe’” (Boellstorff, 2014). Here, the rotten does not necessarily equal ‘spoiled’ — as implicated in the term “link rot”, for example; the process by which hyperlinks break, die and become unavailable. Instead, “rotten” might just as well refer to data that has simply come about in a non-planned and non-organized way, and that may have done so without the direct aid of human minds and hands.

This line of thinking, which recognizes that non-intentionality, chance, and sometimes direct noise is a natural part of any dataset, also coincides with Luciana Parisi’s [4], argument that “*randomness* has become the condition of *programming culture*.” Furthermore, it correlates with Ernst’s (2012) recognition of the fact that technological media actively usher in (and not least stores) autonomous forms of culture and knowledge that do not necessarily follow human logic.

As Parisi [5] explains, algorithms today have the capacity to produce new forms of automated determinations, or what she calls “computational adventure[s]”, that result in new cultural realities. Here, algorithms that manage information are seen as agents who may think and act autonomously upon the world, once their creators have set them in motion [6]. In what follows, I use Lévi-Strauss and Boellstorff’s notion of “rotten” data, and Parisi’s notion of computational randomness as a way to explore the unexpected contents found in The Echo Nest’s giant archive of music metadata. Here, I am not primarily interested in how algorithmic modes of thought may (or may not) affect culture, but rather how they take their own shape and follow their own algorithmic and digital logics.

Music metadata archives & The Echo Nest

As Morris (2012) has described, the history of large-scale databases for music metadata began somewhere around the year of 1993, when the programmer Ti Kan built the first version of a media player called XMDC. As Morris explains, Kan's system for music metadata management soon grew into a separate database called CDDDB (Compact Disc Database). Built, generated, and sustained by music enthusiasts, the CDDDB represents one of the first large-scale, open-source, and user-generated digital archives, and it quickly became the largest collection of music metadata of its time.

However, in the year of 1998, CDDDB was sold to the software company Escient LLC — a move which caused much debate as it essentially turned the collectively controlled database into an appropriated commodity (see Van Buskirk, 2011; Dean, 2004; Morris, 2012). As Morris reports, it did not take long before access to the database started to be sold, and after a series of acquisitions the data collection is now owned by the multimedia corporation Tribune Media Company, who bought it for US\$170 million in December 2013 (Lawler, 2014).

The history of the CDDDB reveals the crucial role of metadata in the formation and growth of digital music as commodities. As Morris (2012) argues, metadata has played a crucial part in “making music behave” in a digital format, since it adds context — and also provide grounds for legal claims to ownership — concerning musical commodities that would otherwise be little more than anonymous streams of data. Importantly, the history of the CDDDB also uncovers the process of metadata starting to emerge as a commodity to be bought and sold in itself; it reveals the process of metadata becoming a new kind of “industrial object” grown out of intimate relations between humans and machines (Hui, 2014).

The Echo Nest entered the music metadata picture in 2005, along with several other competitors occupied with collecting and creating pathways of understanding through massive corpuses of data about music. In particular, the company grew out of the thesis works of two MIT doctoral candidates: Tristan Jehan and Brian Whitman, and its specialty was to provide computer aided assemblages and generations of metadata (rather than expert, or user-generated ones).

Jehan's [7] dissertation was explicitly inspired by the work of Alan Turing and explored the possibility of building “a machine that defines its own creative rules by listening to and learning from musical examples.” His thesis work included the construction of advanced tools of audio analysis — so called *machine listening* tools — with an ability to automatically predict, classify, and make immediate judgments about audio elements of musical pieces. The purpose of these investigations, Jehan explained, was to create a computerized system capable of performing the creative act of composing synthesized music — all by itself. Such an invention, he imagined, could “enable a more intimate listening experience by potentially providing the listeners with precisely the music they want, whenever they want it” [8].

Whitman's (2005) dissertation entitled “Learning the meaning of music”, followed similar trails of thought. Rather than primarily turning to audio analysis however, Whitman focused on computer aided semantic analysis of text to dig deeper into the essence of music. Having a clear predictive element to it, the aim of Whitman's research was partly to create an automatic system that could foresee personal tastes, and provide individually tailored music

recommendations by studying what people say about music online. If Jehan's research was about finding a way to make a computer 'listen' to music to uncover its inner traits and meanings, Whitman's investigations were thus instead an attempt to make a computer 'read' its way to similar results.

Since its foundation The Echo Nest has gone through two large venture capital investment rounds, securing a total injection of \$US24.5 million (The Echo Nest, 2012). The Echo Nest has also received over \$US500,000 in state funding from the U.S. National Science Foundation to develop their "fanalytics" department; a corporate branch that builds tools for "targeted music promotion, demographic and psychographic profiling of artist fan bases, and behaviorally targeted advertising" (The Echo Nest, 2010).

In April 2014, The Echo Nest was acquired by the streaming service Spotify; a deal that resulted in an integration of the two corporations [9], and allegedly set back Spotify at least US\$100 million (Flanagan, 2014). This acquisition reveals how the production and treatment of metadata is central for digital music distribution. As streaming platforms are increasingly doing away with privileging user initiated search functions, and instead favor systems that *push* curated content to its users [10], metadata is no longer serving as an informative side order to the machineries of online music platforms. It is transforming into its main course.

Putting metadata to work

On The Echo Nest's client list, one can find several digital platforms for music (iHeartradio, Rdio, SiriusXM, Deezer, Beats Music, Rhapsody, and of course Spotify), as well as a long range of other businesses, media corporations, and networks for social media (for example BBC, MTV, Twitter, VEVO, Yahoo!, Tumblr, Nokia, Coca Cola, Intel, Microsoft, Reebok, and X-Box). The Echo Nest's collection of metadata has also been embedded and put to work in over 400 digital music applications, including (in The Echo Nests own words) "the #1 Android music player ... and 3 of the top 10 iOS music apps." [11]

In other words, the Echo Nest's archive of music metadata is continuously being commissioned to do work in the world; it is constantly being *consulted*. As it is woven into everyday life through apps and software clients, it exemplifies how the classic, static, and place-bound archive is being transformed into a *moving* digital one [12]. This is not only true in the sense that this is an archive that *captures* movement (online flows of data), and is itself constantly being transformed, updated and adjusted (Parikka, 2012; Røssak, 2010), but also in the sense that it continuously pushes *other* things in motion by uplifting, hiding, and evaluating both people (musicians) and things (sounds). In doing so, The Echo Nests media archive is a dynamic entity that collapses former divisions between archives and everyday life; it facilitates the entanglement of data and culture.

The influence of metadata in contemporary digital music distribution is calling for a need to once again re-think how we conceptualize musical artifacts. Metadata is deeply involved in producing musical experiences and *making music happen* (in similar ways as electronic devices, Internet providers, roadies, and concert ticket salesmen are; see Sterne, 2014; Small, 1998). In doing so, metadata and its managers are erasing distinctions between sounds and

information; divisions between content, and intelligence about contents. Out comes a situation in which a digital file rarely exists in solitude and is instead accompanied and linked together with other types of data in complex, influential, and profound ways.

This process is once again challenging the classic conception (or dogma) that music is something confined in a single musical artifact; the idea that it is an autonomous, independent *work of art*; a “thing” with unique and sovereign boundaries that carries an aura of authenticity and exclusiveness to it (Benjamin, 2008). Instead, taking metadata seriously means recognizing its role in weaving together sounds and information into continuous flows of data — and in extension, to recognize digital music as something processual, porous, entangled, and dynamic.

As we will see, however, this connectivity between sounds and information is often far from frictionless, and rarely seems to stay within the planned roadmaps created for algorithms and software systems. If metadata serves to control and domesticate music (Morris, 2012), the practice of domesticating metadata itself, might still pose challenges for information managers.

APIs as gateways to data

Given the dispersion and influence of The Echo Nest’s archive, it is intriguing to think about the billions of materials that dwell inside this giant database. How is an archive like The Echo Nest’s organized, and how does one make sense out of its contents? From where are its contents pulled to begin with?

The Echo Nest’s application programming interface (or API) provides a base from which to explore these questions. In simplified terms, an API is a structured gateway into a digital source. It allows its users to search, retrieve and implement data from a database or digital platform based on a defined set of rules and queries. The provision of an API is far from unique for The Echo Nest, and the history of APIs have been traced back to the 1980s when new principles for software design were needed to “ensure interoperability between different systems” (Bucher, 2013). A primary function of APIs is that they allow for data and services to become interconnected and built into each other. APIs are thus mediators of data — they allow archives, software and users to mingle [13].

For the purposes of this paper, an experiment was set up in collaboration with Fredrik Palm, system developer at HUMlab, Umeå University. Here, we used The Echo Nest’s API to have a look at their data collection, and were given the permission to use it for research purposes. Simply put, Palm designed an application that made it possible to automatically search, retrieve and store parts of the information that The Echo Nest collects in its database [14]. An external Web site that allowed for the continuous monitoring of the data extraction was also put in place, and equipped with these tools, we set out to monitor The Echo Nest’s generation of metadata concerning 22 different musicians and composers, having origins in 15 different countries during one months time (see [Table 1](#)).

Table 1: List of selected artists + their origin and gender.

Musician(s)	Celebrity range	Origin	Gender
ABBA	Widely known	Sweden	M/F
Johann Sebastian Bach	Widely known	Germany	M
Caetano Veloso	Widely known	Brazil	M
Fela Kuti	Widely known	Nigeria	M
Billie Holiday	Widely known	U.S.	F
Madonna	Widely known	U.S.	F
MC Lyte	Widely known	U.S.	F
Queen	Widely known	England	M
Donna Summer	Widely known	U.S.	F
Fleetwood Mac	Widely known	England/U.S.	F/M
Bob Hund	Locally known	Sweden	M
Vox Vulgaris	Locally known	Sweden	M
Parov Stelar	Locally known	Austria	M
Bariş Manço	Locally known	Turkey	M
Lizzy Mercier Descloux	Locally known	France	F
Jaakko Eino Kalevi	Locally known	Finland	M
Seinabo Sey	Locally known	Sweden	F
Laura Pausini	Locally known	Italy	F
Dorthe Kollo	Locally known	Denmark	F
BABYMETAL	Locally known	Japan	F

Beyoncé	Top charts	U.S.	F
Avicii	Top charts	Sweden	M

Ten of the selected musicians were female, 10 were male, and two consisted of music groups with mixed genders. Ten of the chosen musicians also represented artists and composers who are widely known and have been famous for a long period of time (for example Johann Sebastian Bach and Billie Holiday), while ten of them represented artists who are lesser — and more locally — known (for example the Danish dance band singer Dorthe Kollo, or the Finnish electro musician Jaakko Eino Kalevi). The two remaining artists were represented on several top music charts around the world (Avicii and Beyoncé) [15].

It is important to emphasize that this list of musicians is not intended to be representative of the contemporary global field of artists and composers (for example, European and American musicians are over-represented on the list). All chosen artists were checked and matched with The Echo Nest’s database beforehand — an archive that at the time of this article’s writing (Fall 2015) claimed to “know” over three million artists. In other words, artists who were already identified by The Echo Nest were deliberately chosen, leaving a large portion of the world’s musicians aside. In the end, information concerning the latest updated data entries about the selected musicians was collected once per day by our application between 1 March 2015 and 1 April 2015.

Unlocking The Echo Nest’s archive

The Echo Nest’s music metadata consists of a wide range of different types of materials, and our experiment focused on one of its data types: metadata relating to musicians [16]. More specifically, the following pages will focus on The Echo Nest’s collection of blog posts, which was one of the most frequently updated contents that was pulled into their database. Blogs represent a particularly interesting type of material since this is the kind of data that The Echo Nest use for their textual analysis. Together with reviews, tweets, news articles, and biographies, it is blog posts that underlie The Echo Nest’s efforts to capture Web-based tastes, sentiments, and opinions. The Echo Nest’s collection of blogs may thus be seen as a type of data having been ‘readied’ for algorithms; that is, as information which has been assembled, organized, and labeled for the purpose of algorithmic calculations (Gillespie, 2014). A closer look at blogs thus function as a window into the selection process of The Echo Nest’s analytical system.

In total, our API experiment resulted in securing 1,386 blog posts that The Echo Nest found on the Web [17]. Each of these blog posts were connected to one of the 22 musicians that were chosen for the experiment. Using classic Microsoft Excel software, the posts were analyzed individually according to language and content. In this process, interest was placed on understanding the types of information and knowledge these blogs provide about artists; what kind of data were the company’s Web crawlers programmed to detect? What could this tell us about the proceedings of their curatorial system? Answering such questions, a thematic

approach was applied where all analyzed posts were tagged and divided into to three different categories: (1) blog posts where the musician, band, or artist in question was in focus; (2) blog posts where the musician, band, or artist in question was mentioned; and, (3) what came to be called “rotten” blog posts.

Posts that were tagged “musician in focus” commonly had the relevant artist, composer, or music group in its title, and/or fulfilled the criteria of putting the correct musician(s) in focus throughout its text. A blog post that was connected to BABYMETAL and offered a review of a BABYMETAL concert in Bologna [18], a blog post connected to Barış Manço that centered on one of his live performances on Turkish State TV [19], or a blog post connected to Billie Holiday that announced a planned commemoration of (what would have been) her 100th birthday at the Apollo Theatre in New York [20], were for example placed in this section. Likewise, a post connected to Madonna that discussed the failure of the sales results of her new album in the U.S. [21], or a post connected to Lizzy Mercier Descloux that told the story of one of her old rock bands — Rosa Yemen [22], were tagged as “artist in focus” since they provided wholesome information about the correct musician. In total, about 22 percent of the analyzed blogs were of this kind.

Posts that primarily put other artists, events or topics in focus — but that still made it possible to track down the relevant artist on the Web page by doing a simple word search — were instead labeled “musician mentioned”. For example, a blog post connected to Avicii that fully centered on an interview with one of his former tour partners: the DJ Miss Nine [23], or a post connected to Caetano Veloso that listed the 50 best Brazilian albums of 2014 and briefly mentioned Veloso’s name (even though none of his albums were on the list) [24] would be given an “musician mentioned” tag. Similarly, a post connected to ABBA, written by a man who had just solved a weekly *New York Times* crossword puzzle (where ABBA was part of the quiz) [25], or a post connected to Johann Sebastian Bach that discussed the general music taste of cats (and briefly mentioned that cats apparently “rub against the music speakers” when they hear Bach’s ‘Air on a G String’) [26], were placed in this category, which ultimately made up 68 percent of the analyzed blog entries.

The dominance of posts where artists were only mentioned, indicate that what matters the most in the “digital reputation economy” (Hearn, 2010), is the aggregated presence of a person online; that is, the quantity of *digital* appearances, and not their quality. Web crawlers that sift through the Internet appear to be easy flirts; being mentioned once may be enough to attract their attention. The Echo Nest’s preoccupation with peripheral source materials is indicative of a development where content analysis of trivial textual materials is coming to underlie more of the data turbines that help to traffic music across the Web. Seeing a text about the music tastes of cats as a valid source for information about Johann Sebastian Bach, or approaching a text about solving the *New York Times* crossword puzzle as a source for valuable knowledge about ABBA, suggests that the frivolous is, indeed, a serious matter within contemporary modes of algorithmic expertise.

On the one hand, this might be said to undermine classic forms of musical authority and expertise, as it continues the process of increasingly putting the power of cultural assessment in the hands of the masses. These ‘masses’ are what Gillespie (2014) has called ‘calculated publics’ — a selected group of individuals whose digital behavior and knowledge is extracted, and later re-inserted into the system again in the form of processed data. In such a situation, the role of the musical expert shifts from a (single) human editorial specialist, to the aggregated discussions of large Web populations.

On the other hand, the prevalence of blogs where artists were only briefly mentioned, significantly supports Bowker's (2013) point that we seem to be entering a point in time in which you do not exist unless you are data. While systems like The Echo Nest's may offer new possibilities to move beyond traditional canons (offering wider and different kinds of navigations among musical works), as well as possibilities to transcend traditional identification labels like static genres and stereotypical notions of gender (Cheney-Lippold, 2011), it is still a form of knowledge production that heavily relies on artists having a quantitative presence online. If you're not mentioned on the Web, you're outside the radar of knowledge producing mechanisms, like The Echo Nest.

As it seems, a digital presence matters most if English is used. A closer look at the analyzed posts revealed that even though 15 of the 22 selected musicians had origins in non-English speaking countries, 89 percent of the analyzed posts were written in English. About half of the non-English posts were written in Italian, and pulled from mainly one source: an Italian Web site entitled *soundsblog.it*. The limits of language in The Echo Nest's selection of source materials has significant implications for artists who are active in local (non-English) environments. Regardless if being "seen" by music analytic machineries is considered desirable or not (by musicians and fans alike), the privileging of English "online talk" does indicate that the data powered feeds of online platforms seem to pull listeners towards the English mainstream.

Rotten data

In what remains of this paper, focus will be placed on the last category of blog posts found in The Echo Nest's archive; posts that neither fitted the category "musician in focus" or "musician mentioned". In total, such data represented about 10 percent of the analyzed posts, and this is the data that I propose may be labeled as "rotten." One of the advantages of calling this data rotten (in line with both Boellstorff and Lévi-Strauss) is that it does not impose value on the data. By labeling it rotten, it is not necessarily 'bad', 'spoiled', or 'damaged'. Instead, the term rotten is aimed at saying something about how these materials come about; it is meant to speak to the origins of content, their beginnings as archival objects, and their creation as the result of computational surveillance, detection, and storage.

The concept of the rotted is also meant to reveal how "partial, temporary, and contingent" every understanding of an algorithmic system is (Seaver, 2014). As Seaver (2014) has pointed out, "even when we 'know' an algorithm in broad strokes, we actually know precious little." Calling data rotten is not meant to discern or unveil the inner logics and workings of the algorithms that collected it. Instead, it aims to point out that much of such logics are beyond us (Barocas, *et al.*, 2013; Kockelman, 2013). Rather than thinking about rotten as deviant and flawed, I suggest it might be understood as an integral part of large-scale Web crawling of today. Indeed, it gives evidence of the ways in which noise is incorporated in every instance of communication (Shannon, 1948; Silver, 2012). In particular, rotten links point to the vibrant (and peculiar) agency of software systems and algorithms, with their own tendencies and propensities (Parisi, 2013; Bennett, 2010).

There are four different kinds of rotten data: dead links, orphan posts, name confusions, and re-posts. Not surprisingly, the most common trait among the rotted links had to do with the decay of hyperlinks to specific Web sites. Many of the rotted posts deemed relevant by The Echo Nest, led to Web pages that were no longer operational. These dead links made it impossible to get an overview of their connections to specific musicians. Dead links reveal the relative fragility of the Web (Barone, *et al.*, 2015), and also indicate that possibilities to look into the past and use archives, like The Echo Nest, as an historical resource is limited — both for researchers and software developers alike. The Echo Nest’s database hosted curious archival gaps that pose problems for anyone interested in the past in digital time. The rapid mortality of the Web is one way through which algorithmic systems of knowledge production exert and demonstrate their interpretative power. Since it is often impossible to validate the sources of algorithmic knowledge production, algorithms simply require that we trust them.

Another puzzling type of post appearing in The Echo Nest’s archive were links to blogs where the name of the correct artist could simply not be found anywhere on the page. In short, the contents scraped and provided by The Echo Nest’s API lead to a dead end of orphan blog posts. Neither text nor images on these pages indicated any connection to the related artists, and in many cases the blog posts had little (if anything) to do with music. For instance, Beyoncé was linked to a post about the musician Paloma Ford [27], a blog entry about Kim Kardashian recently dying her hair blond [28], and a post about a man having built a drone replica of the Star Wars Imperial Star Destroyer ship [29], even though Beyoncé’s name could not be found anywhere on the page. Likewise, Billie Holiday was connected to a *Rolling Stone* post about The Doors, Radiohead, and Joan Baez being added to the U.S. National Recording Registry [30], and Madonna to a site that listed “20 New York Dishes You Need To Eat Before You Die” [31], even though I was not able to establish any connection between these posts and the artists in question.

Like dead links, these orphan posts may be the products of a transient and rapidly transforming Web, where associations between people and things are established and erased in a rapid pace. It is, after all, possible that the correct artists were once mentioned in these posts. For example, these sites often had continuously updated newsfeeds where content was quickly removed. At the time of my analysis, however, traces of such previous feeds were gone.

The existence of these posts may possibly also be the result of a distracted Web crawler; an online software surveyor having focused its attention elsewhere, if only for a brief moment in time. Building on the works of Haraway (1987), Bogard (2000) has noted that distraction does indeed appear to have “gone ‘cyborgian’” within contemporary digital media systems. As Bogard suggests, such cyborgian distractions may be traced in surprising scenes of data capture, and perhaps The Echo Nest’s collection of orphan blogs could be seen as such an example.

Confusion over names was yet another common characteristic of the links that I call rotten. Troubles with discerning the right artists from large corpuses of text frequently led to artists being confused with the wrong musicians, wrong celebrities, and sometimes even non-human entities. For understandable reasons, this usually affected composers, artists, and bands with generic names. For instance, ABBA was frequently connected with religious blog posts concerning “Abba” in the Hebrew meaning of the word, where it literally translates to “father.” This, for example, happened when ABBA was connected to a Christian prayer missionary’s blog discussing spiritual hunger [32], and when ABBA was confused with an

excerpt of the Jewish Talmud, quoted in a blog post discussing Orthodox Judaism and rape [33]. Madonna was, for example, also confused with the band Black Honey, and their song entitled “Madonna” [34]. Johann Sebastian Bach had a blog post of the tour dates of the band Iceage associated with his name. One of Iceage’s concerts was to take place at a music venue entitled *Clwb Ifor Bach* in Cardiff [35].

Likewise, the classic rock band Queen was confused with song titles such as “The Queen” by Gentle Giant [36], as well as a user who calls him/herself “Queen By-Tor” and is active on the forum progarchives.com [37]. Queen was also mixed up with the movie *Queen and Country*, a classic drama set in Britain during World War II [38], as well as blog posts concerning several kinds of musical queens, such as Madonna (a.k.a. “the queen of pop”) [39], Nocturnal Sunshine (a.k.a. “the queen of techno”) [40], Mary J. Blige (a.k.a. “the queen of R&B”) [41], Erykah Badu (a.k.a. “the queen of Dallas”) [42], and the Austrian singer and drag queen Conchita Wurst [43]. Queen was also coupled with a news article speculating about Ed Miliband’s likelihood of winning the British elections (and accordingly his chances to visit the Queen of England at Buckingham Palace) [44], a blog post mentioning a rapper from Flushing in Queens, New York [45], and a post from an interior blog displaying US\$300 denim bed linen available for queen and king sized beds [46].

For bands, artists, and composers like ABBA, Queen, Madonna, and Bach, links based on flawed name recognition quickly diluted into much larger pools of data, but for other musicians issues with name recognition had larger effects. For example, data concerning the Danish dance band singer Dorte Kollo — who only had eight blog posts connected to her name in total — was entirely inaccurate and concerned the completely wrong Dorthes. In seven of the cases, Kollo was confused with the Danish author Dorte Nors, and in one instance she was jumbled up with the Danish singer Dorte Gerlach.

The last category of blog entries that were labeled rotten were re-posts, or, links to posts that collected earlier by The Echo Nest’s crawlers. It became possible to detect these re-posts since each new blog entry that is discovered by The Echo Nest is assigned an individual ID number. Comparing ID numbers of posts that were suspected to re-occur revealed that the same kinds of materials were often yanked twice, and sometimes repeatedly, by their systems.

Since each of The Echo Nest’s collected blogs are scanned for descriptive terms (and possibly also for the appearance of other artists’ names that later may be categorized as “similar artists”), it was interesting to note that ABBA became connected to the death metal band Entombed A.D not only once (because they had once recorded their music in the same studio), but twice, since the original post was discovered anew on at least two occasions. This déjà vu moment most likely both increased the measured “hottness” value of ABBA, and curiously also might have strengthened the connection between ABBA and the musically divergent band Entombed A.D.

To summarize, the presence of orphan links, re-posts, dead links, and confusing blog points to some of the problems with ‘automatic’ systems harvesting information from the Web. Their ‘noisiness’ gives evidence of a kind of “random margin” (in lack of a better word to replace the word “error margin”) that appears to mark attempts to detect tastes, sentiments, and preferences online automatically. This is a random margin produced by the difficulties of domesticating automatic systems for information retrieval as well as the unruliness of algorithms and digital source materials that do not always behave as planned. Source materials die, slip out of hand, or are repeatedly re-discovered, leading algorithms to establish

odd connections between people and objects. Placed at the intersection between being humanly ‘untouched’ and humanly ‘managed,’ these rotten links reveal the subconscious of infrastructures being implemented on the Web.

Conclusion

In today’s digital landscape, metadata needs to be understood as a key player in setting the stage for what becomes musical performances. Metadata is not an entity that sits discreetly in the background and assists those who actively search for it. Instead, it is something which is constantly being marshaled to steer, direct, push, and withdraw musical pieces and artistry on digital platforms. Because of these consequences, I argue that it is crucial to take metadata seriously and to explore how it organizes information.

One of the major consequences of the increased reliance on metadata is that it transforms the ways in which digital music objects are conceptualized. In a digital landscape where music is steadily linked and woven together with contextual information, it is no longer viable to think about audio files as autonomous entities. A music file found and enjoyed through the Web is just as much a product of metadata. The actors that deal with metadata — such as The Echo Nest — are central to the process of conflating the boundaries between music and information about music. In doing so, metadata managers are forging new musical materialities and new forms of musical expertise.

As this paper has shown, such forms of expertise appear to be preoccupied with placing trivial, mundane, and sometimes far-fetched source material at the foundation of musical analysis. They also seem to create shadow areas of attention, especially in terms of language. But I would argue that the most interesting ingredient in The Echo Nest’s production of musical know-how is its foundation on surprising, and sometimes outright odd, computationally detected materials from the Web. The presence of rotten data within The Echo Nest’s artist datasets uncovers how algorithmic processes insert new kinds of determination, cultures, and modes of thought into the world. This article has not so much focused on the effects and implications of such algorithmic realities, but instead explored their own internal logics. Algorithms are not simple tools that reproduce information that they consume. Instead they are entities that carry their own cognitive capacities of relating to (and creating) information and knowledge. Randomness seems to be an inevitable part of such computerized agency. The curious types of materials found in The Echo Nest’s database testify to the unpredictable and arbitrary forms of knowledge production that mark large-scale attempts of capturing content and sentiment from the Web.

Writing about the noise and scratches that were once commonly inscribed to old recordings and other media, Ernst decisively (and quite delightfully) once exclaimed that such clatter should be considered the “pure message of the medium”; the roars and voices of machines having been engraved into recordings [\[47\]](#). For media archeologist Ernst, such noises are machinic messages that require our attention; they are a special kind of substratum — or matrix — which is “neither purely human nor purely technological, but literally in between” [\[48\]](#).

In a similar way, the curiosities in The Echo Nest's database represent the inhumanly human whispers of programmed Web crawlers; their distractions (or fascination with orphan links), their hick-ups (or repeatedly new-found Web discoveries), their peculiar association skills (or their name confusion), and their attachment to the deceased (their maintained relations to dead links). One important thing about such rotten content is that it is not discarded matter that has lost its influence on the world. The cognitive abilities of digital architectures are "contagious" (Parisi, 2013) in the sense that their behavior catches and creates ripple effects. In this way, rotten data continues to perform work in the world. It supports, affects, and adjusts continuous identifications and measurements of artistry. Rotten content is thereby not necessarily disqualified to serve as a source of evidence. In the process of identifying aspects that are considered important about artists, rotten blends in with the "cooked" and the "raw" and together forms the basis for the evaluation of musicians.

About the author

Maria Eriksson is a Ph.D. candidate in media and communication studies at the Department of Culture and Media Studies, and an affiliated researcher at HUMlab, Umeå University, Sweden. She has previously received B.A. and M.A. degrees in social anthropology at Stockholm University and is currently part of the interdisciplinary research project "Streaming cultural heritage — Following files in digital music distribution." For more information, visit www.streamingheritage.se.
E-mail: maria [dot] c [dot] eriksson [at] umu [dot] se

Acknowledgments

This research was done as part my involvement in an interdisciplinary research project entitled "Streaming cultural heritage" funded by the Swedish Research Council. I would especially like to thank Pelle Snickars, Patrick Vonderau, Anna Johansson, Johan Jarlbrink, Fredrik Norén, Petter Bengtsson, Vasco Castro, and Robert Eriksson for thoughtful comments to earlier versions of this paper. I also owe a great thanks to Fredrik Palm, who made the experimental parts of the article possible.

Notes

1. However, it should also be noted that a central debate among librarians has been whether or not 'metadata' should be defined as incorporating only digital materials (Vellucci, 2001), or both digital and non-digital contents (Caplan, 2003; Gilliland, 2008; Gill, 2008). The former viewpoint is one that disqualifies clay tablets and other pre-modern forms of structured information collections from being 'metadata'. In this paper however, I adopt a broader conceptualization of metadata, and follow the strand of thinkers who do not see a necessity in separating digital 'metadata' from other kinds of data.

2. boyd and Crawford, 2012, p. 663.

3. Lévi-Strauss, 2013, p. 41.

4. Parisi, 2013, p. i.

5. Parisi, 2013, p. x.

6. I want to emphasize, however, that by calling these elements rotten (and thus recognizing the autonomy of algorithmic forms of knowledge production), I am not arguing that humans are stripped of accountability with regards to the workings of systems like The Echo Nest. On the contrary, the ability of these systems to *do stuff*, and exert power on the world, lies in the hands of their creators and administrators. Ultimately, it is us who give algorithms the power to *think* and *behave*. Once allowed the freedom to act, however, algorithms do generate autonomous cultural realities (Parisi, 2013), and it is these curious realities that the article explores.

7. Jehan, 2005, p. 23.

8. Jehan, 2005, p. 25.

9. Tight collaborations between The Echo Nest and Sportily had, however, existed for a longer period of time. For example, The Echo Nest has managed Spotify's radio function since 2009; see The Echo Nest press release, dated 18 May 2009, at <http://the.echonest.com/pressreleases/echo-nest-teams-spotify/>.

10. See, for example, the Spotify press release of 20 June 2015, at <https://news.spotify.com/us/2015/05/20/say-hello-to-the-most-entertaining-spotify-ever/>.

11. To see a full list of the apps that use The Echo Nest's analytics, visit <http://the.echonest.com/showcase/>, accessed 15 October 2015.

12. Parikka, 2012, p. 120.

13. There are, however, two conflicting descriptions of APIs that circulate. From an industry perspective, APIs have been framed as creative technical solutions that “open” data to a wider range of actors and allow for new possibilities for revenue (O'Reilly, 2002). At the launch of The Echo Nest's developer API in 2009, it was for example argued that this move should be understood as an example of the goodness of “relinquishing control” over technology (Hyde, 2009). On the other hand, other (more critical) voices have pointed out that APIs aid “capitalistic enclosure” by putting in place “synergistic membranes with prescribed circuits that lead less and less to the global Web outside their online properties” (Milberry and Anderson, 2009, p. 394). By weaving together and creating dependencies between platforms and users, APIs are part of business strategies that opens up — at the same time as it strengthens — the power and influence of the actors who own and control them. APIs have therefore been described as entities that specify “protocolized relations between data, software, and hardware” (Dijck, 2013, p. 32), that mediate the types of analyses that can be made from data (Burgess and Bruns, 2012), and that embody a certain ‘politics of data’ (Bucher, 2013). Taken together, these circumstances call for a need to understand APIs as objects of study in themselves, rather than as neutral tools of content delivery.

14. Put in more technical terms, Palm scheduled and executed a PHP-script that ran every day at 7.30 am. The extracted data was first stored in a JSON-format and organized into a database that connected the data to each artist and the date/time of capture. Since The Echo Nest's API only allows a restricted amount of requests for data per second, the PHP-script had a built-in delay in each request for data, so that the API's terms of use would not be violated. This means that data concerning different artists was never captured at the exact same time, but rather in a serial fashion. On the whole, however, the difference in scheduled data capture never exceeded one hour.

15. A new record release by Madonna (who was originally thought of as "widely known") did, however, push her up the charts and thus might have made it suitable to place her in the category of being on the "top charts" as well.

16. For example, this means that The Echo Nest's analysis of sounds (their 'machine listening' section), and their analysis of user behaviors and tastes (what they call their "fanalytics" department) was left to the side. Similarly, the analysis did not focus on images, reviews, biographies, and news articles that The Echo Nest also monitors and scrapes from the Web on a daily basis. Neither does the article take a closer look at The Echo Nest's measurements of artistry, that is, their collection of "top terms" (a list of the most commonly used words to describe a musician/band), "discovery" values (an estimation of "unexpected artist popularity"), "hottness" values (an estimation of how much attention is being given to a musician/band on the Web), or "familiarity" measurements (estimations of how "well known" an musician/band is).

17. It should be noted, however, that our extraction of The Echo Nest's data revealed some significant differences in terms of the amounts of data being added to each composer, musician, or band name. For instance, an artist like Beyoncé had a wide variety of new materials added to her name every day. However, links to local artists like Dorthe Kollo turned out to be several years old. Likewise, for an artist like Beyoncé — who had a total amount of over 50,000 data points connected to her name — the information detected by our application only represented a small fraction the data about her. For others, like Dorthe Kollo, we could examine all of her personal data, since it amounted to a much smaller quantity.

18. See <http://www.soundsblog.it/post/341392/babymetal-in-concerto-a-giugno-a-bologna-tutte-le-informazioni>, accessed 20 October 2015.

19. See <http://www.undomondo.com/post/109197204969/bar-s-manco-scheherazade-sehrazat-turkish>, accessed 20 October 2015.

20. See <http://www.billboard.com/articles/news/6487723/apollo-theater-to-honor-billie-holiday-on-her-100th-birthday>, accessed 20 October 2015.

21. See <http://www.tonspion.de/neues/neueste/5515661>, accessed 20 October 2015.

22. See <http://altcitizen.com/listen-rosa-yemen-rosa-vertov-1980/>, accessed 20 October 2015.

23. See <http://swedeandsour.tumblr.com/post/112676817769/miss-nine-february-22-2015-interview>, accessed 20 October 2015.

24. See <http://miojoindie.com.br/os-50-melhores-discos-nacionais-de-2014-10-01/>, accessed 20 October 20 2015.
25. See <http://rexwordpuzzle.blogspot.com/2015/03/poet-cassady-who-was-friend-of-jack.html>, accessed 20 October 20 2015.
26. See http://www.refinery29.com/2015/03/83275/cats-musical-real-cats-respond-video?source_utm=feed&utm_medium=rss, accessed 20 October 20 2015.
27. <http://2dopeboyz.com/2015/03/05/paloma-ford-jada/>.
28. <http://www.thehollywoodgossip.com/2015/03/kim-kardashian-did-she-go-blonde-to-steal-the-spotlight-from-ken/>.
29. See <http://mashable.com/2015/03/25/imperial-star-destroyer-drone/>, accessed 20 October 20 2015.
30. See <http://www.rollingstone.com/music/news/the-doors-radiohead-joan-baez-added-to-national-recording-registry-20150325>, accessed 20 October 20 2015.
31. See <http://www.refinery29.com/where-to-eat-nyc#slide-15>, accessed 20 October 2015.
32. See <https://soulpants.wordpress.com/2015/02/24/hunger-is-meant-to-produce-growth/>, accessed 20 October 2015.
33. See <http://theunorthodoxjew.blogspot.com/2015/03/reading-oral-law-today-forces-jews-to.html>, accessed 20 October 20 2015.
34. See, <http://www.sos-music.co.uk/2015/02/black-honey-madonna.html>, accessed 20 October 2015.
35. See, <http://pitchfork.com/news/58696-iceage-announce-world-tour/>, accessed 20 October 2015.
36. See <http://www.progarchives.com/Review.asp?id=1372768>, accessed 20 October 2015.
37. See <http://www.progarchives.com/Review.asp?id=1373214>, accessed 20 October 2015.
38. See http://laist.com/2015/02/28/queen_and_country_review.php, accessed 20 October 2015.
39. See <http://telegraph.feedsportal.com/c/32726/f/568374/s/43e5b80f/sc/17/l/0L0Stelegraph0O0Cculture0C114421280CMadonna0EMiley0ECyrus0Eis0Ejust0Eexploring0Eher0Esexuality0Bhtml/story01.htm>, accessed 20 October 2015.
40. See <http://www.repeatbutton.com/electronic/nocturnal-sunshine-take-me-there/>, accessed 20 October 2015.

41. See <http://hiphop-n-more.com/2015/03/video-mary-j-blige-doubt/>, accessed 20 October 2015.
42. See http://blogs.dallasobserver.com/dc9/2015/03/the_best_concerts_in_dallas_this_weekend_327_329.php, accessed 20 October 2015.
43. See <http://www.billboard.com/node/6502188>, accessed 20 October 2015.
44. See <http://www.telegraph.co.uk/news/politics/labour/11441694/The-tuition-fees-farce-proves-Labour-isnt-ready-to-govern.html>, accessed 20 October 2015.
45. See <http://www.complex.com/music/2015/03/best-action-bronson-songs/>, accessed 20 October 2015.
46. See <http://www.coolhunting.com/buy/scottie-store-chambray-bed-sheets>, accessed 20 October 2015.
47. Ernst, 2012, p. 69.
48. Ernst, 2012, p. 70.

References

- Solon Barocas, Sophie Hood, and Malte Ziewitz, 2013. "Governing algorithms: A provocation piece," essay prepared for *Governing algorithms: A conference on computation, automation, and control* (29 March), at <http://governingalgorithms.org/resources/provocation-piece/>, accessed 20 October 2015.
- Francine Barone, David Zeitlyn, and Viktor Mayer-Shönberger, 2015. "Learning from failure: The case of the disappearing Web site," *First Monday*, volume 20, number 5, at <http://firstmonday.org/article/view/5852/4456>, accessed 23 May 2016.
doi: <http://dx.doi.org/10.5210/fm.v20i5.5852>, accessed 23 May 2016.
- Walter Benjamin, 2008. "The work of art in the age of its technological reproducibility: Second version," In: Michael W. Jennings, Brigid Doherty, and Thomas Y. Levin (editors). *The work of art in the age of its technological reproducibility, and other writings on media*. Cambridge, Mass.: Belknap Press of Harvard University Press, pp. 19–56.
- Jane Bennett, 2010. *Vibrant matter: A political ecology of things*. Durham, N.C.: Duke University Press.
- David M. Berry, 2011. "The computational turn: Thinking about the digital humanities," *Culture Machine*, volume 12, at <http://www.culturemachine.net/index.php/cm/article/viewArticle/440>, accessed 23 May 2016.

Tom Boellstorff, 2014. "Making big data, in theory," *First Monday*, volume 18, number 10, at <http://firstmonday.org/article/view/4869/3750>, accessed 23 May 2016.
doi: <http://dx.doi.org/10.5210/fm.v18i10.4869>, accessed 23 May 2016.

William Bogard, 2000. "Distraction and digital culture," *C-Theory* (5 October), at <http://www.ctheory.net/articles.aspx?id=131>, accessed 20 October 2015.

Geoffrey C. Bowker, 2013. "Data flakes: An afterword to "*Raw data*" is an oxymoron," In: Lisa Gitelman (editor). "*Raw data*" is an oxymoron. Cambridge, Mass.: MIT Press, pp. 167–171.

danah boyd and Kate Crawford, 2012. "Critical questions for big data," *Information, Communication & Society*, volume 15, number 5, pp. 662–679.
doi: <http://dx.doi.org/10.1080/1369118X.2012.678878>, accessed 23 May 2016.

danah boyd and Kate Crawford, 2011. "Six provocations for big data" (21 September), at http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1926431, accessed 20 October 2015.

Taina Bucher, 2013. "Objects of intense feeling: The case of the Twitter API," *Computational Culture*, number 3 (16 November), at <http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api>, accessed 23 May 2016.

Jean Burgess and Axel Bruns, 2012. "Twitter archives and the challenges of 'big social data' for media and communication research," *M/C Journal*, volume 15, number 5, at <http://journal.media-culture.org.au/index.php/mcjjournal/article/viewArticle/561>, accessed 23 May 2016.

Priscilla Caplan, 2003. *Metadata fundamentals for all librarians*. Chicago: American Library Association.

John Cheney-Lippold, 2011. "A new algorithmic identity: Soft biopolitics and the modulation of control," *Theory, Culture & Society*, volume 28, number 6, pp. 164–181.
doi: <http://dx.doi.org/10.1177/0263276411424420>, accessed 23 May 2016.

Kate Crawford, 2013. "The hidden biases in big data," *Harvard Business Review* (1 April), at <https://hbr.org/2013/04/the-hidden-biases-in-big-data>, accessed 20 October 2015.

Katie Dean, 2004. "The house that music fans built," *Wired* (7 July), at <http://archive.wired.com/entertainment/music/news/2004/07/64033>, accessed 20 October 2015.

José van Dijck, 2013. *The culture of connectivity: A critical history of social media*. Oxford: Oxford University Press.

Wolfgang Ernst, 2012. *Digital memory and the archive*. Minneapolis: University of Minnesota Press.

Andrew Flanagan, 2014. "Spotify's Ken Parks on an IPO, the Echo Nest Purchase (Q&A)," *Billboard* (11 March) at <http://www.billboard.com/biz/articles/news/digital-and->

mobile/5930388/spotify-ken-parks-on-an-ipo-the-echo-nest-purchase-qa, accessed 20 October 2015.

Tony Gill, 2008. "Metadata and the Web," In: Murtha Baca (editor). *Introduction to metadata*. Second edition. Los Angeles: Getty Research Institute, and at https://getty.edu/research/publications/electronic_publications/intrometadata/metadata.pdf, accessed 23 May 2016.

Tarleton Gillespie, 2014. "The relevance of algorithms," In: Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (editors). *Media technologies: Essays on communication, materiality, and society*. Cambridge, Mass.: MIT Press.
doi: <http://dx.doi.org/10.7551/mitpress/9780262525374.003.0009>, accessed 23 May 2016.

Tarleton Gillespie, 2012. "Can an algorithm be wrong?" *Limn*, number 2, at <http://limn.it/can-an-algorithm-be-wrong/>, accessed 23 May 2016.

Anne J. Gilliland, 2008. "Setting the stage," In: Murtha Baca (editor). *Introduction to metadata*. Los Angeles: Getty Research Institute, and at http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf, accessed 23 May 2016.

Lisa Gitelman (editor), 2013. *"Raw data" is an oxymoron*. Cambridge, Mass.: MIT Press.

Lisa Gitelman and Virginia Jackson, 2013. "Introduction," In: Lisa Gitelman (editor), 2013. *"Raw data" is an oxymoron*. Cambridge, Mass.: MIT Press, pp. 1–14.

Donna Haraway, 1987. "A manifesto for cyborgs: Science, technology, and socialist feminism in the 1980s," *Australian Feminist Studies*, volume 2, number 4, pp. 1–42.
doi: <http://dx.doi.org/10.1080/08164649.1987.9961538>, accessed 23 May 2016.

Alison Hearn, 2010. "Structuring feeling: Web 2.0, online ranking and rating, and the digital 'reputation' economy," *Ephemera*, volume 10, number 3, pp. 421–438, at <http://www.ephemerajournal.org/contribution/structuring-feeling-web-20-online-ranking-and-rating-and-digital-'reputation'-economy>, accessed 23 May 2016.

Yuk Hui, 2014. "Metadata," In: *Critical keywords for the digital humanities*, at <http://cdckeywords.leuphana.com/metadata> (not found), accessed 20 October 2015.

Ben Hyde, 2009. "Demand the surprises" (5 January), at <http://enthusiasm.cozy.org/archives/2009/01/demand-the-surprises>, accessed 20 October 2015.

Tristan Jehan, 2005. "Creating music by listening," Ph.D. dissertation, Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology, at http://web.media.mit.edu/~tristan/phd/pdf/Tristan_PhD_MIT.pdf, accessed 23 May 2016.

Paul Kockelman, 2013. "The anthropology of an equation: Sieves, spam filters, agentive algorithms, and ontologies of transformation," *Hau: Journal of Ethnographic Theory*, volume 3, number 3, pp. 33–61, at <http://www.haujournal.org/index.php/hau/article/view/hau3.3.003>, accessed 23 May 2016.

Markus Krajewski, 2011. *Paper machines: About cards & catalogs, 1548–1929*. Cambridge, Mass.: MIT Press.

Ryan Lawler, 2014. “Tribune has closed its acquisition of Gracenote ... Here’s what comes next,” *TechCrunch* (3 February) at <http://techcrunch.com/2014/02/03/tribune-has-closed-its-acquisition-of-gracenote-heres-what-comes-next/>, accessed 20 October 2015.

Claude Lévi-Strauss, 2013. “The culinary triangle,” In: Carole Counihan and Penny Van Esterik (editors). *Food and culture: A reader*. Third edition. New York: Routledge, pp. 40–47.

Claude Lévi-Strauss, 1969. *The raw and the cooked*. Translated by John and Doreen Weightman. New York: Harper amp; Row.

Astrid Mager, 2012. “Algorithmic ideology,” *Information, Communication & Society*, volume 15, number 5, pp. 769–787.
doi: <http://dx.doi.org/10.1080/1369118X.2012.676056>, accessed 23 May 2016.

Lev Manovich, 2011. “Trending: The promises and the challenges of big social data” (28 April), at <http://manovich.net/content/04-projects/067-trending-the-promises-and-the-challenges-of-big-social-data/64-article-2011.pdf>, accessed 20 October 2015.

Viktor Mayer-Schönberger and Kenneth Cukier, 2013. *Big data: A revolution that will transform how we live, work, and think*. London: John Murray.

Kate Milberry and Steve Anderson, 2009. “Open sourcing our way to an online commons: Contesting corporate impermeability in the new media ecology,” *Journal of Communication Inquiry*, volume 33, number 4, pp. 393–412.
doi: <http://dx.doi.org/10.1177/0196859909340349>, accessed 23 May 2016.

Jeremy W. Morris, 2012. “Making music behave: Metadata and the digital music commodity,” *New Media & Society*, volume 14, number 5, pp. 850–866.
doi: <http://dx.doi.org/10.1177/1461444811430645>, accessed 23 May 2016.

Tim O’Reilly, 2002. “Inventing the future” (9 April), at <http://archive.oreilly.com/pub/a/network/2002/04/09/future.html>, accessed 20 October 2015.

Jussi Parikka, 2012. *What is media archaeology?* Cambridge: Polity Press.

Luciana Parisi, 2013. *Contagious architecture: Computation, aesthetics, and space*. Cambridge, Mass.: MIT Press.

Cornelius Puschmann and Jean Burgess, 2014. “Metaphors of big data,” *International Journal of Communication*, volume 8, pp. 1,690–1,709, and at <http://ijoc.org/index.php/ijoc/article/view/2169>, accessed 23 May 2016.

Eivind Røssak, 2010. *The archive in motion: New conceptions of the archive in contemporary thought and new media practices*. Oslo: Novus Press.

Nick Seaver, 2013. "Knowing algorithms," *Media In Transition* 8 (Cambridge, Mass., April), at <http://static1.squarespace.com/static/55eb004ee4b0518639d59d9b/t/55ece1bfe4b030b2e8302e1e/1441587647177/seaverMit8.pdf>, accessed 20 October 2015.

Claude Shannon, 1948. "A mathematical theory of communication," *Bell System Technical Journal*, volume 27, number 3, pp. 379–423.
doi: <http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x>, accessed 23 May 2016.

Nate Silver, 2012. *The signal and the noise: Why so many predictions fail — but some don't*. New York: Penguin Press.

Christopher Small, 1998. *Musicking: The meanings of performing and listening*. Hanover, N.H.: University Press of New England.

Pelle Snickars, 2014. "More music is better music" (16 December), at <http://pellesnickars.se/2014/12/more-music-is-better-music/>, accessed 20 October 2015.

Spotify, 2015. "Say hello to the most entertaining Spotify ever" (20 May), at <https://news.spotify.com/us/2015/05/20/say-hello-to-the-most-entertaining-spotify-ever/>, accessed 20 October 2015.

Jonathan Sterne, 2014. "There is no music industry," *Media Industries Journal*, volume 1, number 1, pp. 50–55, and at <http://www.mediaindustriesjournal.org/index.php/mij/article/view/17/26>, accessed 23 May 2016.

The Echo Nest, 2012. "The Echo Nest raises over \$17 million in new financing to bring big data to music" (12 July), at <http://the.echonest.com/pressreleases/echo-nest-raises-over-17-million-new-financing-bring-big-data-m/>, accessed 20 October 2015.

The Echo Nest, 2010. "The Echo Nest receives National Science Foundation Phase IIB Small Business Innovation Research (SBIR) Grant" (4 March), at <http://the.echonest.com/pressreleases/echo-nest-receives-national-science-foundation-phase-iib-small-/>, accessed 20 October 2015.

The Echo Nest, 2009. "The Echo Nest teams up with Spotify" (18 May), at <http://the.echonest.com/pressreleases/echo-nest-teams-spotify/>, accessed 20 October 2015.

Eliot Van Buskirk, 2006. "Gracenote defends its evolution," *Wired* (13 June) at http://www.wired.com/2006/11/gracenote_defen/, accessed 20 October 2015.

Sherry L. Vellucci, 2001. "Music metadata and authority control in an international context," *Notes*, volume 57, number 3, pp. 541–554.
doi: <http://dx.doi.org/10.1353/not.2001.0063>, accessed 23 May 2016.

Brian Whitman, 2005. "Learning the meaning of music," Ph.D. dissertation, School of Architecture and Planning, Program in Media Arts and Sciences, Massachusetts Institute of Technology, at <http://hdl.handle.net/1721.1/32500>, accessed 23 May 2016.

Editorial history

Received 6 November 2015; revised 15 May 2016; accepted 17 May 2016.

This paper is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

Close reading big data: The Echo Nest and the production of (rotten) music metadata
by Maria Eriksson.

First Monday, Volume 21, Number 7 - 4 July 2016

<https://journals.uic.edu/ojs/index.php/fm/rt/prINTERfriendly/6303/5530>

doi: <http://dx.doi.org/10.5210/fm.v21i7.6303>