

# Managing by Data: Algorithmic Categories and Organizing<sup>1</sup>

To appear in Organization Studies

<https://journals.sagepub.com/doi/abs/10.1177/0170840620934062>

**Cristina Alaimo**

LUISS Guido Carli University, Italy

**Jannis Kallinikos**

LUISS Guido Carli University, Italy

London School of Economics, UK

---

<sup>1</sup> **Corresponding author:**

Cristina Alaimo LUISS Guido Carli University, Viale Romania 32, 00197 Roma, Italy.

Email: [calaimo@luiss.it](mailto:calaimo@luiss.it)

# Managing by Data: Algorithmic Categories and Organizing

**Cristina Alaimo**

LUISS Guido Carli University, Italy

**Jannis Kallinikos**

LUISS Guido Carli University, Italy

London School of Economics, UK

Data and data management techniques increasingly permeate organizations and the contexts in which they are embedded. We conduct an empirical investigation of Last.fm, an online music discovery platform, with a view to unpacking the work of data and algorithms in the process of categorization. Drawing on Eleanor Rosch and her colleagues, we link the making of categories with the construction of basic objects that function as key filters or registers for perceiving and organizing the world and interacting with it. In contexts such as the ones we have studied, basic objects are made out of data rather than expert or community-based knowledge. In such settings, basic objects work as pervasive reality filters and as the entities on which other organizational objects and categories are built. As they diffuse, such objects and the categories they instantiate become naturalized, increasingly reconfiguring the social order of organizations and their environments as a data order. Once key organizational activities such as the making of objects and categorizing are rearranged by data and algorithms, organizations can no longer be framed as separate from the technologies they deploy.

**Key words:** algorithms, basic objects, categories, classification, data, organizations, technology

## Introduction

Categories are pervasive human constructs and categorization a widespread daily and institutional practice. We view categories and categorization schemes as culturally embedded (community-based or expert-based) modes of ordering reality which organizations variously draw upon to design, conduct and evaluate their operations (Bowker & Star, 1999; Douglas, 1986). The foundational function that categories perform in organizations recurs in several widely influential works in the history of organization studies. Perrow (1986) views categories as performing a role similar to stereotypes whereby the perceptions of particular objects and events are assigned into larger and uniform groups that are coped with through the use of standardized programmes (see also Scott, Ruef, Mendel &

Caronna, 2000; Weick, 1976). Mintzberg (1979), in particular, links professional bureaucracy to a *pigeonholing* process, whereby clients are assigned to specific groups that are treated by invoking particular and standardized procedures. Pigeonholing is largely a process of developing and maintaining categories. The role categories play in organizations and fields has more widely been investigated by reference to the development of expert knowledge and the establishment and maintenance of professional categorization schemes and industry practices (Bowker & Star, 1999; Desrosières, 1998; Lena & Peterson, 2008; Star, 1999; Timmermans & Berg, 1997). In helping mediate the world experts confront, categories are linked to structural attributes of organizations such as the division of work and the communication and authority lines in which expert work is embedded (Mintzberg, 1979; Scott et al., 2000).

The pervasive nature of categorization schemes grants them a nearly infrastructural presence in organizations. Categorization schemes often extend beyond single organizations and cut across a field (such as medical or legal knowledge), providing the cognitive grid (e.g. medical diseases, crime types) by means of which communities, work groups and experts deal with the realities they confront. Cast in this light, categories work as taxonomic systems and rather seldom in isolation. In this regard, our focus on the role categories play in organizations and fields differs substantially from recent research on the dynamics of market categories as largely labelling practices that contribute to the construction and stabilization of products and industries (i.e. Negro, Hannan & Rao, 2011; Hannan, Pólos & Carroll, 2007; Rosa, Porac, Runser-Spanjol & Saxon, 1999). Ours is a study on the construction of facts and organizational objects via the medium of algorithmic categories and the data they encode. We pursue this line of research with a keen awareness of the fact that community and expert categorization schemes are currently being reframed by the profound cultural and organizational involvement of data and algorithms (Alaimo & Kallinikos, 2017; Bechmann & Bowker, 2019; Faraj, Pachidi & Sayegh, 2018; Flyverbom, 2019; von Krogh, 2018).

Our research links data, and the means by which they are produced and made sense of within and across organizations, to the process of categorization. Online forms of organizing, in particular, such as those exemplified by digital platforms (e.g. Gawer, 2009) or social media platforms (Alaimo & Kallinikos, 2019), are variously based on clustering various types of data with a view to finding new or more efficient ways to offer value to users and other stakeholders. Facebook, LinkedIn and TripAdvisor, for instance, are all typical examples of such forms of organizing. Most of platform operations entail one or another form of data clustering and categorization achieved via the deployment of data-management techniques and widespread organizational use of algorithms (Beverungen, Beyes & Conrad, 2019; Rosenblat, 2018). Placed against this backdrop, we ask the following questions: a) how do data, algorithms and data-based systems interfere with the creation of organizational facts, categories and objects in the fields in which they operate? b) How do they shape the paths along which organizations interact with their environments? Ultimately, c) what are the broader implications of these developments for new practices and forms of organizing?

We deal with these questions by focusing on the intermingling of data, the technologies that support the production and diffusion of data and the AI-related methods by which data are made to matter. For reasons of brevity and convenience, we refer to this intermingling as the apparatus of data-technologies-algorithms and explore its interlacing with categorization and the process of organizing. Our focus on the work of this apparatus dissociates categorization from its micro-foundations and directs attention to the cultural and structural properties of organizations and the technological systems by which organizational facts are reproduced. Our objective is to give an account of data practices and categorization that reflects the wider forces at work and captures their embedment into larger cultural, social and technological contexts (Abdelnour, Hasselbladh & Kallinikos, 2017; Douglas, 1986; Knorr-Cetina, 1999).

We empirically address the work of this apparatus through the investigation of *Last.fm*, one of the oldest digital platforms dedicated to music discovery and an early adopter of the data-driven approach to music taste. We study the process of data categorization and the role categories play in framing and organizing the reality that platform stakeholders confront. Our findings indicate that platform operations converge around certain types of data which are first engineered and subsequently deployed as the basis for categorizing music taste. A decisive step in this process is the making of basic objects out of data (e.g. artist names) which furnish the means for the development of higher-order categories such as similar artists that support the organizational objective of music discovery. Basic objects often work as boundary objects. It is through basic objects that the organization has over the years negotiated the redefinition of many traditional music operations such as identifying music genres, classifying music and recommending artists and tracks. Drawing from our findings, we analyse the process of organizing and unpack the ways it is linked to the apparatus of data-technology-algorithms.

The rest of the paper is structured as follows. The next two sections deal with the theory of categorization and the role categorization schemes play in social fields and organizations. We briefly review a range of literatures and provide our account of how categories are involved in the conceptual scaffolding of organizational work. We then move on to presenting the empirical study of *Last.fm* and our findings. In the analysis section we assess the theoretical value of our arguments for the process of categorization and further discuss the relevance of our findings for organizations and fields. Granted the pervasive nature of the operations we document, we conclude that it is not any longer fruitful to treat technology as an exogenous force, separate from the organizational operations into which it is embedded. Organizations and technology co-constitute one another and have accordingly to be studied in tandem (see also, Beyes, Holt & Pias, 2020; Kallinikos, Marton & Hasselbladh, 2013).

## **Categories and Cognition**

Categorization is a widespread institutional practice but also a fundamental human operation, part of the unspoken fabric of cognitive habits and conventions that govern social interaction and community life. At this level, categories furnish the epistemic registers, the filters through which individuals and social groups sort out the variety of experience into classes of objects or events on the basis of attributes they share with one another. According to Eleanor Rosch, whose ideas have had a profound impact on the contemporary understanding of categories, categorization is a primal social activity on the basis of which singular, *non-identical* objects or events (stimuli in their own terminology) are treated as largely *equivalent* and, accordingly, sorted out into similar groups (Rosch, 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976).

Categories are often linked to broader categorization or classification schemes. It is a widespread and important attribute of categorization schemes to entail categories that relate to one another in terms of inclusion, whereby lower-level categories are subsumed under higher ones which possess additional classes of members or objects (e.g. humans and mammals or finance and service organizations). The more inclusive a category, the higher its level of abstraction and the fewer the properties or attributes on the basis of which it is derived (Rosch et al., 1976). Abstract categories enable selectivity and precision often at the expense of reality purchase. By the same token, the less inclusive a category is the wider the range of properties on the basis of which it is derived.

One of the key insights stemming from the work of Rosch and her colleagues is that social interaction evolves around what they call *basic objects*; that is, the perception of equivalent groups of things in ways that achieve a viable and useful balance between abstraction (e.g. low world content) and concreteness (too many details). Ordinary, daily interactions are intertwined with the cultural establishment of basic objects (chair, table, car) abstract enough to allow individuals and communities to achieve an economy of perception (e.g. bypassing the differences between individual types of tables, chairs or cars) while dealing with real and observable world entities. It is in this regard that basic

objects strike a balance between the informational content they convey for social actors and the tangible or context-embedded objects they refer to. More correctly, a basic object is the highest level of abstraction (e.g. a table, a car), at which concrete references to real-world entities can still be established. In this regard, basic objects perform an information-rich function while maintaining proximity to tangible reality which superordinate categories (e.g. furniture, vehicle) tend to attenuate. In the words of Rosch and colleagues: “The basic level of classification, the primary level at which cuts are made in the environment, appears to result from the combination of these two principles; the basic categorization is the most general and inclusive level at which categories can delineate real-world correlational structures” (Rosch et al., 1976, p. 384).

Cast in this light, categorization delivers the basic blocks of reality and, while culturally embedded, it is conditioned by the divisions and segments of the natural and social world, what many people understand as objective reality. Such a broadly realist view of categorization stands at a variance with a diffused conventionalism that assumes that the world can be divided up in any way we want (for a relevant debate see, e.g. Douglas, 1986; Hacking, 2001). Yet, such realism by no means rules out the influence which cultural constructs and conventions may have upon the ways by which reality is perceived and ordered. In being a basic social activity, categorization is inevitably linked to community life, natural language (e.g. nouns and verbs) and perception (Eco, 2000; Lakoff, 1987).

In fact, it has been one of Rosch’s great social science contributions to demonstrate that the cultural nature of categorization entails the mapping of the perception of objects to a *prototype* on the basis of resemblances that exhibit a *graded intensity or structure* (Rosch & Mervis, 1975). On this view, categorization is not an either/or activity but a graded exercise. The epistemic but also practice implications of this seductively simple, ultimately Wittgensteinian, idea of family resemblance are that objects and the categories to which they belong lack clear-cut boundaries. Such condition confers

certain freedom of action concerning how “cuts are made in the environment” and categories and basic objects can be perceived and established (Bowker & Star, 1999).

## **Categorization, Organizations and Fields**

The constructs of basic objects, prototype and family resemblance are useful conceptual devices for understanding the work of expert groups and the establishment of organizational processes and functions. Research on medical classifications and the ways categorization is used by health care professionals is a case in point (see e.g. Bowker & Star, 1999). Similarly, research conducted in other fields such as natural science (Star & Ruhleder, 1996), higher education (Espeland & Sauder, 2007) and the financial sector (MacKenzie, 2006, 2017) shows the significance of categories in the infrastructuring of field perceptions and practices. Categories provide distinctions (“cuts in the environment”) that work as knowledge standards and action protocols that infrastructure experience and help coordinate experts and practice communities (Bowker & Star, 1999; Kornberger, Pflueger & Mouritsen, 2017).

A central theme which runs through the study of categorization is the perennial issue concerning how ordinary forms of perception and social interaction mediated by categories (community-derived basic objects) interact with formal modes of classification and standardization propagated by professions and organizations (Bowker & Star, 1999; Knorr-Cetina, 1999; Timmermans & Epstein, 2010). Terms such as local versus global or structure versus action (Kallinikos & Hasselbladh, 2009; Kornberger et al., 2017; Star & Ruhleder, 1996) are indicative of the tension between the distinctions and categories of communal life (i.e. the interaction order) and those of institutions and fields (the structural order) (Berger & Luckman, 1967; Goffman, 1974).

“Large-scale industrial processes are their own institutions. They cannot be embedded in the patterns of local community control”, Mary Douglas claimed (1986, p. 108) explaining the evolution of classifying practices in the history of textile and wine production. Her historical analysis of these two



industries shows that the exigencies of production for large, often global, markets required more abstract attributes on the basis of which products were classified. Product categories such as Cashmere for textiles or Bordeaux for wines belonging to early modern, geography-bound classes were refashioned and products classified on the basis of more abstract and context-free attributes such as type of wool or grape. Such categories and the more generic industry practices of classifying products on the basis of their underlying materials were tied to the shifting exigencies of production that rendered geographical origin superfluous. We interpret Mary Douglas's account as indicating that the specialized actions characteristic of mature institutional fields, sooner or later, bring out distinctions and categorizations that move away, complement and occasionally undermine the basic objects (here geographical denominations of product categories) characteristic of an industry or field (see also Labatut, Aggeri & Girard, 2012).

Understanding the tensions that the infrastructural work of categorization carries within organizations and across fields is essential to assess the impact of current data-based categorization and how it differs from previous technological regimes. The change brought about by techniques of data management and algorithmic ordering is undeniably linked to wider transformations in technologies (hardware, software) and institutional fields within which such change is embedded (Ekbjær & Evans, 2009; Kallinikos et al., 2013; Zuboff, 1988). Yet, it is of utmost importance to investigate how techniques of data management and algorithmic ordering embody new and context-free ways of segmenting and organizing field processes and objects that reframe both community-based and field-specific practices of categorization and the classification schemes to which they are linked.

Current computing technologies are powerful means for capturing and analysing facts in the form of data and information. The sheer scale of the operations they are called upon to monitor combines with the formal and analytic origins of computer science to disseminate modes of clustering and aggregating data that sidestep central principles of categorization embedded in communal or expert-

based practices (Kallinikos & Hasselbladh, 2009; Zuboff, 1988). Computer-based techniques of data ordering and management rely on *data tables* and *fields* which work as minute registration filters at a far more granular level than those of established categories and community-based knowledge (Rosenfeld & Morville, 2002). An important consequence of this is the bottom-up construction of bigger data groups or categories out of the primary data provided by data tables and fields. The identification and recording of facts through data fields (which is often a production of facts) do not any longer need to pass through the medium of socially-embedded or cultural categories and the categorization schemes they belong. For instance, people can currently be profiled or categorized on the basis of tracking and assembling minute individual online actions, instrumented as clicks and browser-overs. The ways such techniques assemble facts substantially differ from attributing individuals to classes such as age or gender, geographical location, income or lifestyle (Zwick & Dholakia, 2004). In many contemporary contexts of institutional life, data-based techniques are used to complement and, in some cases, transform and bypass established professional categories and practices (Flyverbom & Murray, 2018). Personalized medicine and learning analytics represent typical examples (Hamburg & Collins, 2010; Siemens & Long, 2011).

## **Methodology**

Our empirical investigation is a single case study of Last.fm that focuses on the mechanisms and operations through which the platform automatically sorts out online music experience by aggregating data, deriving from the listening preferences and other behaviours of its users. Last.fm has been selected because it represents a typical example of data-driven organization that has been explicitly established to disrupt existing expert categories in the field of music consumption. We conduct our empirical research as a first step toward knowledge development and, eventually, theory building (Eisenhardt, 1989; Yin, 2009). Our objective is to integrate the empirical findings to a broader framework that shows how the apparatus of data-technologies-algorithms works in the environment of online platforms and how it sustains different forms of organizing.

Our data collection and analysis draw on digital methods (e.g. Rogers, 2013). Relying mostly on online archival records and publicly available online documents, we gathered information from different sources to enhance validity through data triangulation and the iterative cross-checking of recurring themes that emerged out of the empirical material as we moved from data collection to the extraction of preliminary findings. To this end, we collected all the blog entries of Last.fm (from its inception in May 2007 to its termination in January 2014), a total of 210 documents that we analysed and coded using thematic coding, that is, the identification of semantically connected set of ideas in the empirical material. We also gathered Last.fm Application Programming Interface (API) documentation and snapshots of the Last.fm platform at different points in time between October 2002 and December 2018 to cover all the major events of the platform (i.e. platform redesign, changes in platform functionalities). Furthermore, we collected and analysed the Last.fm's official forum data. The Last.fm official forum is an online community made by different discussion groups formed by Last.fm users. We extracted forum threads from Last.fm and selected those with 10 or more responses for our analysis. This produced 113,030 entries which we queried using the keywords derived from the first round of thematic coding, performed on the 210 blog entries. In the context of digital platforms user fora constitute places of discussion and negotiation between the organization, end users and developers. To complement our sources and further validate our analysis we also collected external documents and data. These included articles from The Guardian (106 documents) and blog entries from Techmeme, one of the most reputed aggregators of technology blogs (127 documents) which we analysed using a second round of thematic coding, and discussion threads from Stackoverflow, a well-known forum for developers (1,189 documents) which we used to further validate some of the themes and findings from previous stages of our data analysis. Table 1 below summarises our sources of evidence and the type of analysis we performed on each one of them.

**Table 1:** Data sources and analysis of Last.fm

Data sources	Number of entries collected	Data analysis
Last.fm blog (from May 2007 to January 2014)	210 documents	Thematic coding (first round): (i) we analysed all the entries (ii) found themes and (iii) extracted a set of keywords for further analysis and triangulation
Snapshot of Last.fm platform at different points in time from The Internet Archive Wayback Machine service of archive.org	223 snapshots of Last.fm “News” “Announcements” and “Updates” sections (between 2002 and 2017) API documentation and 291 snapshots of build.last.fm (developer website) (between 2008 and 2018)	
Last.fm forum (official and unofficial)	113,030 entries (collected between 2015 and 2016)	Keywords and theme validation: (i) we run several searches using the keywords extracted from the first round of thematic coding to validate, enrich, split or merge themes
The Guardian and Techmeme (from 2006 to 2014)	106 and 127 documents respectively	Thematic coding (second round): (i) we analysed all the entries (ii) found themes and (iii) extracted a set of keywords for triangulation
Stackoverflow	1,189 documents (collected between 2015 and 2016)	Keywords and themes validation: (i) we run several searches using the keywords extracted from the first round of thematic coding to validate, enrich, split or merge themes

## The Social Music Revolution of Last.fm

Last.fm, one of the oldest organizations for online music discovery, was founded in 2002, as a radio and social music discovery platform, offering music streaming and online radio services, social networking features and music personal recommendations. The platform was originally established with the aim of disrupting traditional music consumption and advancing an alternative music experience based on actual user listening behaviour, user interactions and a lively online community life. Due to adverse economic performance, this ideal had to undergo considerable adaptation over the years. A turning point in the history of the platform was the discontinuation of its streaming services in 2009 which was followed by the termination of online radio services in 2014. In the aftermath of these significant events, declining user participation brought about the winding down of several of the platform’s social media features and functionalities in 2015. Today, the core of Last.fm revolves almost exclusively around music recommendation services.

Despite the failure to fully realize the revolution it once envisioned, Last.fm has nonetheless had a lasting impact on online music consumption. Many of the technologies that empower today’s

streaming platforms, music data service providers and music data-based repositories were originally developed by Last.fm and were closely linked to the platform's original ideal of a socially empowered music experience. The two fundamental assumptions that underpinned this ideal were that music consumption is a largely a social behaviour and that a data-driven approach is a better way to discover music taste than traditional expert-based approaches. The first of these assumptions, a vision of a 'social music revolution' materialized over the years through the development of various social features and user profiles built around music taste that were mostly discontinued in 2015. The second assumption was realized through the development of Audioscrobbler and its music recommender system, a technology that currently forms the core of Last.fm's operations. Audioscrobbler is a technology that constructs a detailed user listening profile by tracking user's listening behaviour across a number of online music media players, Internet radio stations, connected devices and other streaming platforms such as Spotify and YouTube. By downloading Audioscrobbler's plugins, users submit listening data whenever they play tracks online.

In what follows we describe in some detail the fundamental operations on which music recommendations are based. The aim of the narrative is to link music recommendations with the construction of basic objects and categories and disclose the mechanisms, data operations and entities involved in the construction and diffusion of such recommendations.

### ***Tracking user listening behaviour***

To offer personalized music recommendations, Last.fm needs to gather, store, order and categorize music data. Differently from previous systems of categorizing based on expert or industry classifications of artists and music styles, Last.fm operates by gathering data produced by the real-time music listening behaviour of users which it clusters and categorizes bottom-up, using big data and machine learning techniques. To personalize its music recommendations, Last.fm uses an item-based collaborative filtering recommender system which it extends and qualifies with a tagging system (Aggarwal

2016). Item-based collaborative filtering is one of the oldest and most widely used recommendation systems. Collaborative filtering, as the name indicates, filters data from the behaviour of large groups to recommend relevant items to individual users. The system embeds the assumption that relying on real-time data of listening habits from a large music community instead of individual experts is more effective in finding new and relevant ways of categorizing music and filtering recommendations.

By listening to music on any online music player application, computer, or portable device with an Audioscrobbler plugin or extension, users automatically submit their listening choices to Last.fm. Scrobbling creates and transmits to Last.fm a playback-event every time a user listens to a track. Playback events, in turn, generate playcount data which are the counting of how many times a listening event occurs. Playcount data are the main data entity of the system, connecting an individual user to playback events and building the individual and collective history of user listening behaviour. Playcount data are deployed to generate personalized charts and recommendations.

As with many other data tracking technologies, there is no playcount data out there which Last.fm records or tracks. Rather, the platform creates a playcount every time it receives an event that meets specific design requirements and fits a certain pre-determined data format. Put differently, tracking listening data is an organizing process that requires a number of rules and parameters which define all the characteristics the event needs to have in order to produce playcount data. For instance, a playback event needs to contain specific metadata such as artist name, track name, and timestamp. Only if the signal contains such parameters is it recognized as playback event and encoded into listening data (playcount)<sup>2</sup>. The careful crafting of how listening is encoded into data reflects a long

---

<sup>2</sup> To do so, track.scrobble (the web service method for adding track plays) assigns a set of optional parameters to meet systems' requirements such as for instance, album names, whether the track is being chosen directly by users (or a recommender system), track numbers on albums, album artists (if they differ from track artists) and durations of tracks. Also, the submission of listening data works only if: (i) the track is longer than thirty seconds and (ii) the track has been played for at least half of its duration or for four minutes, whichever occurs earlier. In the case in which any of the aforementioned conditions are not met the track is not recognized as a playback-event and is thus not recorded in the system's database. In such cases, the behaviour basically does not exist. If this happens consistently for a particular artist, the artist will not enter in Last.fm's music database and its music recommendations. In Last.fm fora, users have pointed out that some

process of negotiation between the organization and its users and developers which entails more than just the making of inscription rules. Once recognized as listening data, playcounts, in fact, need to pass through numerous filters which are applied to standardize the heterogeneous data gathered from the 600 connected applications and devices<sup>3</sup>.

Last.fm uses artist names (which are data submitted with playback events) as data buckets or pigeonholes. Artist names represent a decisive intermediate passage between the organization's making of data out of selected and encoded events (playcounts) and the subsequent creation of wider and algorithmically-derived categories (similar artists). In contexts such as the one Last.fm represents, artist names are no more than data objects created *ad hoc* by the system to contain the information it needs to automate the computation of recommendations. They become the key objects that function as placeholders within which the system stores the playcounts and other data (metadata) it extracts from the track-file. Although an artist name may exist already in the music field, the organization moulds it into an object which sustains its own data-driven approach to music listening.

Effectively, artist names already exist both offline as names of (real) artists and online as metadata or descriptions of music tracks. Yet, because there is no agreed-upon standard or institutional music identifier to name digitized music, the same music track can have different descriptions (different artist-names). Out of an arguably uncontested entity such as the name of an artist, digitized music has created a multitude of entities (the many different versions of the same artist name such as nicknames, wrong spellings, etc.). Machines cannot yet identify misspellings as different versions of the same name and thus the Last.fm's system automatically creates a new object out of any misspelling of a

---

genres of music, such as grindcore, usually contain tracks that are shorter than 30 seconds and are therefore excluded by audioscrobbler rules.

<sup>3</sup> Sometimes the filter applies to artist-name data, other times to time; in either case listening data that do not meet these conditions are ignored. For instance, last.fm discourages third party developers from determining a track's metadata from its filename, this means that an artist-name that appears to be a filename will be ignored (filtered out) by the system. As some users discuss on the forum, this cuts off actual bands such as the Japanese Moe-pop band MOSAIC.WAV, which name looks like a filename.

given artist's name. The system, for instance, produces as many objects as the different spellings of Louis Armstrong name, where "Louis Armstrong", "Armstrong Louis", "L. Armstrong" each creates an object in the system.

```
$ ./query ../mp3/Pink\ Floyd/The\ wall\ \ (disc\ 1\
Fingerprint ID: 527153
--
Pink Floyd - Another Brick in the Wall (Part 1)
Pink Floyd - Another Brick In The Wall (Part 1)
Pink Floyd - Another Brick In The Wall Part 1
Pink Floyd - Another Brick in the Wall, Par
Pink Floyd - Another Brick in the Wall, Pt. 1
Pink Floyd - The Wall (PartI)
Pink Floyd - Another brick in the wall (p..
Unknown - unknown

Or another example:

$ ./query ../mp3/Guns\ N\ \ Roses\ Use\ Your\ Illus
Fingerprint ID: 1211395
--
Guns N' Roses - Knockin' on Heaven's Door
Guns N' Roses - Knockin' On Heaven's Door
Guns N' Roses - Knockin On Heavens Door
```

**Figure 1.** Example of differently spelled tracks from Last.fm blog, see <http://blog.last.fm/2007/09/10/fingerprinting-update>

This seemingly trivial problem creates data inconsistencies whose scale has always been one of the most challenging computational problems for the system. Last.fm has, over the years, implemented different strategies to select the correct version of an entry: user vote (so called community-based solution of flagging incorrect entries), manual data cleansing and cleansing robots. Here is what a Last.fm engineer said on the issue.

Until recently, the system had no way of identifying variation in spelling of artist and tracks, which led to many duplicate pages for the same artist or track. With millions of scrobbles coming in everyday, it doesn't take a genius to figure out that you will soon have a big metadata mess on your hand<sup>4</sup>.

In 2009 Last.fm implemented *Autocorrection*, an automated system that identifies incorrect artist names and maps them to their correct version. Albeit automated, the task remains hard to conclusively address and is further complicated by the fact that new (correct and incorrect) artist names are constantly ingested to the system. This adds a real-time exponential complexity to an already quite intensive computational task<sup>5</sup> which requires constant updates from internal developers and continuous feedback from users.

---

<sup>4</sup> From Last.fm Blog. See, <http://blog.last.fm/2009/01/29/closing-in-on-clean-metadata-artist-and-track-spelling-auto-correction-is-here>

<sup>5</sup> This entails, for instance, all the (other) metadata associated with incorrect artist entries together with strings of listening patterns that is computationally complex to redirect to correct artist names.



### *From objects to similarities*

A further step in the production of music personal recommendations is the establishment of similarity between artist names. Similarity in this context is not constructed on the basis of the intrinsic attributes of the artist-items grouped. In the system, artist names are basic objects created by counting and clustering user plays. In most essential respects, they are data objects. All that is known of an artist in these online contexts is how many times its name is associated with a user playback. It is on the basis of this link between user playback and artist name that the system establishes similarity between artists. Similarity measures to what degree two or more artists tend to be listened together by two or more users over time. The system assumes that artists that tend to be listened by the same users have something in common, it therefore uses the listening patterns of these users to rank artists and compute suggestions.

Once similarity between artist objects is established, similar artists are put into clusters or, as the jargon goes, neighbourhoods. Such clusters or neighbourhoods are categories of similar artists that work as groups of predictors (see e.g. Aggarwal, 2016). To rank how much similar two artists are within a group of already similar artists, the system will select a measure of similarity (see Alaimo & Kallinikos, 2020). Recommendations are essentially predictions cast in the form of probabilities. To put it simply, a prediction to suggest Ella Fitzgerald will be computed on the basis of the ranking of her nearest neighbours (i.e. Louis Armstrong, Billy Holiday, etc.).

There are a number of limitations to this formal approach to categorization. Similarity in music discovery systems cannot distinguish genres of music, attributes or specific features of artists or bands. Given that similarity is just a score, based solely on listening data, it may very well determine two artists belonging to different genres as similar to one another and group them into the same neighbourhood. The only solution to this problem would seem to be to gather more data or to implement a

hybrid approach to recommendation by adding a different recommender system (Alaimo & Kallinikos, 2020). Not surprisingly, Last.fm has blended its collaborative filtering by implementing a user-generated tag system. Users can tag by attributing any label to any track. Labels are keywords and can be genres, names, years, but also personal labels, based on experience or use such as for instance moods, occasion (i.e. dinner or party) memories (i.e. summer 2018) and so on. Tags seek to provide to the system the missing embedment into a meaningful structure of relations which the counting of listening data lacks. Tags were envisioned to offer an additional dimension to Last.fm's music discovery system, as the statement by a Last.fm engineer below illustrates:

[T]his diversity and eclectic view of the musical landscape is what Last.fm is all about. We don't live in a cookie cutter world of hackneyed generic labels for music. Most music sites include the standard dozen genres (pop, rock, urban, etc.) and that doesn't adequately describe the diversity of music out there. Our tags system encourages the weird and the wonderful, the micro-communities and new scenes that are springing up as fast as new, independent bands are formed<sup>6</sup>.

The results, however, proved to be a mixed blessing. On the one hand, tagging delivered a staggering variety of labels that mixes mostly traditional genres and general attributes with functional value such as "Saturday night" or subjective feelings such as "moody". These miscellaneous labelling practices (tags) certainly brought in novelty and, effectively, Last.fm was one of the first platforms to experiment with suggestions based on its tags with its Discovery web app<sup>7</sup>. On the other hand, the cumulated four million tags have left gaps in identifying novel music characteristics. The tags turned out to be either too personal or altogether predictable. "Although you can make up any tag you like, we noticed that in practice most people use tags that describe genre, or closely related things such as the era, or nationality of an artist", noted an engineer from the *Music Information Retrieval* team at Last.fm<sup>8</sup>. Additionally, the massive scale reached by tags and their intrinsic aggregate value required the development and implementation of new automated methods of analysis, a task that eventually turned out to be a big challenge for Last.fm.

---

<sup>6</sup> From Last.fm Blog. See, <http://blog.last.fm/2011/12/15/announcing-lastfm-discover>

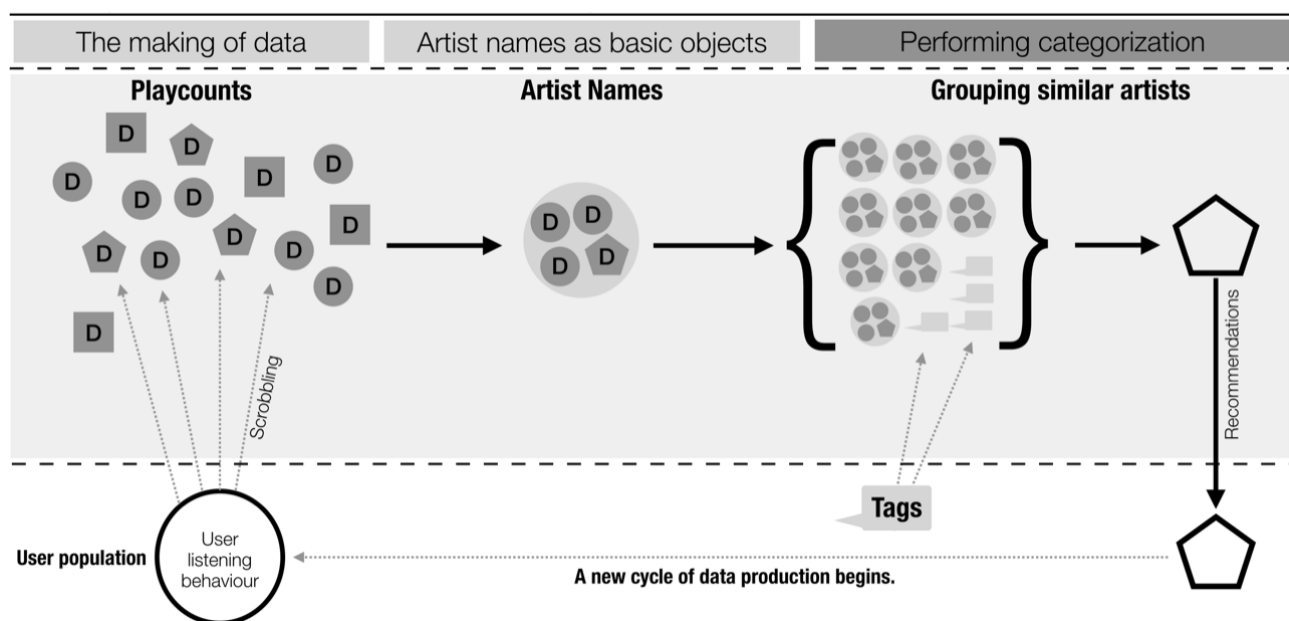
<sup>7</sup> From Last.fm Blog. See, <http://blog.last.fm/2011/12/15/announcing-lastfm-discover>

<sup>8</sup> From Last.fm Blog. See, <http://blog.last.fm/2012/08/30/how-are-you-feeling-today>

The last step of this music recommendation cycle is the computation of prediction. Albeit merely an output of a complex organizational process which involves all the passages we have illustrated above, prediction is the ultimate purpose of recommender technologies. The system is designed to produce a form of knowledge that is not descriptive but predictive and, in the final analysis, performative. The service the organization provides, the design of the system, the encoding of events, the making of basic data objects and the algorithmic assembly of data to categories are constantly optimized to achieve efficient prediction outputs in a massive scale.

## Analysis

We described above the operations of data making (playback events and playcounts) of Last.fm and the establishment of artist names as the stepping stone on the basis of which other categories such as similar artists are built and recommendations advanced. The transition from playcounts via artist names and similar artists to music recommendations describes the flow of data through several stages of elaboration and indicates the basic nature of the organizational work performed by Last.fm. Meanwhile Figure 2 below illustrates the process flow, Table 2 further systematizes these ideas.



**Figure 2:** The work of the data-technologies-algorithms apparatus

Table 2 below depicts the entities we link with the creation of categories; that is, artist names *qua* basic objects, similar artists and popular artists or trends. Each of the categories corresponds to major system operations and key requirements, rules and functions that support them. On the right side of the table, we indicate the general sources of evidence that support our findings.

**Table 2.** Categories, system operations, requirements and functions

Categories	Major system operations	Requirements and functions	Empirical Sources
Artist names	Defining events Tracking and filtering listening events Counting listening events (playcounts) Aggregating playcounts into artist names	Audioscrobbler technologies and extensions (i.e. plugins, APIs) Establishing functional rules and requirements as to what constitutes listening events Defining playcounts as data Implementing Autocorrection System	API documentation Last.fm official Blog Snapshots of Last.fm and uild.last.fm platforms Last.fm official Forum
Similar artists	Measuring similarity Ranking similar artists Implementing tagging and using tagging data to disambiguate similar artists	Item-Item algorithm and hybrid recommender system technologies Platform user participation (Tagging) Real time personalized recommendations	Collaborative filtering publications External documents Last.fm official Blog Last.fm official Forum
Time-bound contextual categories (i.e. Artists for you, Tags to explore, Tracks you need to hear, Blast from the past, Hot right now, etc.)	Ingesting latest playcounts Updating nearest neighbours Incorporating user feedback Updating recommendation	Time stamps of playcounts Feedbacks from users	External documents Last.fm official Blog Snapshots of Last.fm platform Last.fm official Forum

Over the next three subsections, we analyse and elaborate the details of this process. We approach categorization and the organizational implications of categorizing with a view to unpacking the data, conditions and operations that pervade Last.fm. It is important to restate as clearly as possible that we do not study how certain actors or groups (e.g. engineers, managers or entrepreneurs) might have shaped Last.fm. Both the approach we adopt and the empirical evidence we reported above is about what we call the apparatus of technologies-data-algorithms that underlies the process of categorization and furnishes the foundation upon which Last.fm is erected as organization. What we document in this paper is what Knorr-Cetina (1999) refers to in the context of her work as *machineries of knowing*, which in our case translates to the technological forces, cultural perceptions and structural conditions that together constitute the matrix out of which organizations such as Last.fm are made. Such

a perspective may complement actor-oriented explanations of social affairs yet differs from them in the sense of focusing upon processes, forces and operations that extend beyond the micro-foundations of organizational behaviour (see e.g. Abdelnour et al., 2017). Here is how Knorr-Cetina (1999, p. 9) posits the issue:

The action-theory framework offers little purchase for establishing the patterns on which various actions converge and which they instantiate and dynamical extend. It shines the analytic torch upon the strategies and interests and interactional accomplishments of individuals and sometimes groups. While this yields important insights into how agents generate and negotiate certain outcomes, it offers no dividends on the machineries of knowing in which these agents play a part.

### **The making of data**

Playcounts are the key data inputs through which the listening behaviour of users is registered, rendered legible and, ultimately, computable. The making of data that can serve as essential inputs for further operations requires some upfront organizational decisions, which are then embedded in the system. To implement such decisions further requires addressing a series of technical problems which we described in some detail in the preceding pages. Playcount is the count of a user's plays that pass the filters, rules and parameters of Audioscrobbler. Data produced this way are therefore marks or inscriptions of user listening behaviour as this is engineered and subsequently tracked by the system. Playcount is the elementary and fundamental data unit of the system. Via the playcount, the listening of music undergoes a subtle but decisive transformation: it becomes a formal operation made of discreet data, engineered in specific formats and put together to build a computable object.

Data are often, and rather mistakenly, equated to facts out there (Bailey, Leonardi & Barley, 2012; Kallinikos, 1999). Yet, the connection of data to facts is very complex, often insidious. In organizations such as Last.fm, the format of data is produced by the functional prerequisites of the systems in place (i.e. the Audioscrobbler with its music discovery functionalities). As we showed earlier, data are the outcome of a number of contingent processes and organizational negotiations (Jones, Blackwell, Prince, Meakins, Simpson & Vuylsteke, 2019) rather than the straightforward mapping of an

area of reality out there, as the listening behaviour may suggest at first sight. If they represent facts, data often do so in a rather narrow sense. In some cases, such as those of simulation, data may entail strong correspondences or structural affinities to a reality out there (Bailey et al., 2012; Knorr-Cetina, 1999). More often than not, the specific formats of data and the formal language used to record them considerably determine the particular event that is recorded, stored and sorted out (Alaimo & Kallinikos, 2019; Bowker and Star, 1999; Flyverbom & Murray, 2018; Gitelman, 2013). An event is inscribed into data only because it fits the format of the digital token and possess functional attributes such as metadata which are important to its interoperability and computability.

As such, playcounts have no value or meaning on their own. They are made valuable thanks to their volume and the function they perform in the broader system of knowing set up by the organization. Playcounts make listening behaviour countable and possible to enter to a series of permutations with other data. Making data thus entails an act of selection, interpretation and encoding that is at the same time an important organizational decision that conditions subsequent organizational operations, decisions or tasks. Last.fm's system works on the basis of the organizational interpretation of what constitutes a music event (playback events) in online platform settings and, accordingly, negotiates data rules and parameters with users, developers and other stakeholders (e.g. music labels). At the same time, these negotiations are heavily limited by a number of other constrains such as the kind of technology employed, its functional prerequisites and other technical and industrial standards prevailing in the field.

### **Artist names as basic objects**

Playcounts are aggregated into artist names. Artist names operate as basic objects, in the ways defined earlier in this paper, reducing the variability of the world (here playcounts) yet delivering entities concrete enough to aid perception and action (Rosch, 1975; Rosch et al., 1976). As basic objects, artist names allow the organization to perform and coordinate its main activity –that is, the encoding

of music listening behaviour into data— with little or no concern for being representative of entrenched social practices linked to cultural processes of categorization. From its very beginning, the platform's mission has been to innovate online music consumption by relying on automated methods of data management. The making of artist names out of playcounts does not draw from cultural music conventions (i.e. real artist names). It rather provides the system with the scalability and standardization of the data it needs to run accurate statistical parsing and offer recommendations of music out of user listening behaviour data.

Cast in this light, the engineering of basic objects in online settings presents some important differences to the socio-cultural making of basic objects. The balance, in particular, between abstraction and concreteness that is key in coordinating action and communication within and across organizations and fields (Bowker & Star, 1999; Rosch & Marvin, 1975; Rosch et al., 1976) is perturbed and shifted towards abstraction or, perhaps, more correctly, formality. The processes we describe displace the trade-off between concreteness and abstraction underlying basic objects with a vaguely recognizable cultural entity (artist name) that serves the principles of scalability, standardization and computational efficiency, and the cascade of automated operations that enact these principles (Couldry & Mejias, 2019). The flipside of having a music discovery system based on a data object such as artist name with vague socio-cultural roots and no real social context is the production of a babel tower, where it is no longer possible to distinguish a real artist name from a fabrication (Boast et al., 2007). Up until 2009, Last.fm ran with the idea that massive volumes of artist names and the technology already in place would end up with the correct artist name. After different attempts at dealing with misspelled names and inaccurate entries with community flagging and other means, Last.fm implemented the autocorrection system, the automated mapping of non-standard artist names into correct ones. Algorithmic categorization does not admit conflicting rationalities and the only way of solving an issue caused by automation is by implementing further automation (Arthur, 2009; Ciborra, 2009; Kallinikos et al., 2013).

Artist names as basic objects provide the foundation upon which Last.fm derives higher-order categories, based mostly on similarity but also on other time-contingent measures such as popularity and trending. But artist names also operate as boundary objects of a very specific type. In many ways, artist names are the meeting point between Last.fm, users, developers and the platform ecosystem in which Audioscrobbler operates. The knowledge artist names embed allows and constrains the coordination of organizational activity across multiple contexts. Organic boundary objects are both abstract and concrete and have both local and shared meaning which need to be constantly negotiated and balanced out to maintain coherence across intersecting communities (Bowker & Star, 1999; Star, 2010). Artist names however are engineered boundary objects. While they enable massive collaboration across organizational and field boundaries, they at the same time lack the plasticity that would allow them to adapt and speak to different communities of practice. In the case of Last.fm this led to the proliferation of rules and parameters governing their production and use and ended up excluding data sources and membership categories (certain genres of music that do not fit the parameters of data ingestion, external developer practices, etc.) and triggered protests from users and developers. Ultimately, the rigid nature and formalism of data boundary objects affected music recommendation.

### **Categorization and user behaviour**

Artist names qua basic objects provide the basic material out of which higher-order categories such as similar artists, popular artists and other time-bound clusters of artist names such as daily, weekly or monthly trends are derived. Similarity, in particular, is an axial principle and the backbone of all recommender systems that populate the online world of platform organizations. Categories of this sort activate the potential meaning of playcounts and represent a crucial technological step through which entities are related to one another via recommendations.



These recommender systems operate by sorting data objects into similar groups on the basis of the computation of patterns of linked occurrences. Louis Armstrong and Ella Fitzgerald are deemed similar because they are listened (play counted) by the same users over time. In this regard, the activity music recommendation systems perform is functionally equivalent to categorization processes, insofar as they both treat non-identical objects or events as largely equivalent and sort them out, accordingly, into similar groups. Yet, the means and premises by which they achieve these goals are vastly different. In settings such as Last.fm, higher-order categories are the outcome of clustering and aggregating the objects and events such settings bring into being on the basis of formal operations rather than via resemblances or other intrinsic or practice-related affinities.

Once artist names are created and filtered by the autocorrection system, they need to be assigned into similar groups and mapped into artist similarity networks. In this process, similarity is defined as the step by step procedure of calculating the distance between two items (artists) according to how much the rating patterns (listening is unary rating) of two or more users agree. The category of similarity is thus established by computing, not by experience, cognition or knowledge. The results obtained are groups of artists that are made visible to users as similar and, importantly, acted upon as similar. Because of the similar artists category, the system is able to produce a number of derived ephemeral categories such as “Artists for you”, “Tags to explore”, “Tracks you need to hear”, “Blast from the past”, etc. which direct user listening behaviour and further update modes of online music consumption across the online music field. It is in this sense that recommender systems operate in a functionally equivalent mode to traditional categorization. Recommender systems deliver larger cognitive clusters, such as similar artists, yet they do so in ways that differ remarkably from how culturally-embedded categorization schemes mediate the perception of reality.

## **Discussion**

In what follows, we place the discussion of our findings into the wider context of categorization and organization. We first review how data intensive environments, such as the one Last.fm represents, change the process and organizational relevance of categories and categorization schemes. We then move on to consider the implications of our ideas for the current understanding of technology, organizing and organizations.

### **Categorization and Organization**

Institutionally-based classifications inevitably develop, Mary Douglas (1986) claimed, beyond community-based categorization schemes. The rising complexity of production (technical, material, economic) brings the establishment of new ways of labelling and classifying to accommodate the matrix of operations underlying this complexity. The path from community to expert or industry-based classifications is, of course, seldom linear and exhibits considerable empirical diversity. Douglas's claim is nonetheless a good approximation of how the complexity of production and the advances in knowledge that underpin that complexity are linked to new schemes for ordering and categorizing work in organizations and across fields.

Our own findings indicate that these ideas are, *mutatis mutandis*, applicable in the case of Last.fm and, perhaps, more widely to how music experience (and consumption) is organized in online settings. Platform organizations in particular, orchestrate the participation of highly dispersed and atomized populations of users which they seek to maintain and enlarge. At the two endpoints of the process we have analysed stand the rendition of user listening habits into data and the making of recommendations that steadily connect back to users. Both these fundamental operations require categorization of listening data to basic objects such as artist names and the derivation of higher-order categories such as similar artists, processes that entail a long and hidden datawork at a remove from the platform interface. The steady production of recommendations renders the atomized user platform populations as a sort of surrogate community, whereby users are made similar to one another on the basis of a

computed similarity (Alaimo & Kallinikos, 2019). Current experience beyond Last.fm indicates that recommendations increasingly become native in online worlds, and algorithmic categories, such as similar artists, *naturalized* (Bowker & Star, 1999; Douglas, 1986).

Cast in this light, personalized recommendations make visible similar artists and music genres derived *ad hoc* while holding obscure the work of the apparatus, that is, the long journey of formal operations by which similarity is produced in platform organizations. Once re-contextualized and re-socialized, these categories emerge as a common currency that effectively coordinates user interaction online and the organization's encounter with users and other stakeholders (e.g. developers, other platforms, artists). Categories of similar artists are navigated by users and naturalized by their listening behaviour which, in turn, reinforces further algorithmic categorization and aids system learning and optimization. In this regard, categorization redefines experience and long-entrenched cultural habits (Couldry & Mejias, 2019). The apparatus of data-technologies-algorithms in which these developments are embedded operates by sorting things out on behalf of users. Its computational outputs mingle, develop adjacently and, occasionally, take the place of existing socio-cognitive structures and cultural patterns by facilitating the processing of available categories and shaping the emergence of new ones.

Our investigation points out how these processes often require larger time scales to manifest, over which several tensions and conflicts are settled or addressed. Our analysis illustrates how these tensions and negotiations are considerably framed by technological functionalities and developments, the path-dependencies of the digital music field and, more generally, the performativity and self-reference of automation, whereby technologies and technological operations steadily beget new ones (Arthur, 2009; Kallinikos et al., 2013). Once this apparatus is at work, its operations become black-boxed and inaccessible to end users and, to a certain degree, developers and organizational members. The basic objects of categorization such as artist names embed machine readable knowledge that

facilitate automation but make the objects and the process of categorization non-readable to humans. In this regard, algorithmic categorization does not, and cannot, rely on existing socio-cultural contexts while its operations remain at a remove from and opaque to human inspection, understanding and control. After all, the entire apparatus of data-technologies-algorithms is a technical response to the low commensurability, permutability and computability of cultural objects and categories (Alaimo & Kallinikos, 2017; Espeland & Sauder, 2007) which the apparatus reconstructs as data-based operations.

Paraphrasing Knorr-Cetina (1999, p. 40), we can claim that under the online conditions we have outlined in this paper, the social order is reconfigured as a data order. The idea is not entirely new and is encountered across the social sciences (e.g. Bailey et al., 2012; Borgmann, 1999; Foucault, 1970). In the field of organization studies, it can be traced back to Zuboff's (1988) pioneering research on the computerization of work. Computerization, Zuboff foresaw, reconstitutes the order of organizations in the form of data tokens that bring about the far-reaching textualization of work. This electronic text, as she called it, is *comprehensive* and *systemic*, exhibiting considerable *depth* and existing independently of *space* and *time*. This text Zuboff claimed has *no author* in the conventional sense. It is just the patchwork of vaguely-connected individual acts structured and mediated through automation and the power matrix from which it derives (1988, pp. 179-181).

Never before have these ideas been as relevant as they are today. Recasting them in the current context of the web and the online environments of platform organizations certainly entails some important modifications (see e.g. Faraj et al., 2018; Flyverbom, 2019; Kornberger et al., 2017; Monteiro & Parmiggiani, 2019; Von Krogh, 2018). Our own research updates and extends these poignant observations by showing how the apparatus of data-technologies-algorithms encroaches upon and restructures the cultural exchanges of social actors. Organizations such as Last.fm become possible thanks to the codification of cultural habits and the remaking of music preferences by means of data

production processes and the algorithmic categorization schemes they entail. On this account, technology emerges not simply as a mechanical force but also as a pervasive medium (Beverungen et al., 2019) that reorganizes the human sensorium (categories) and the ways social agents perceive and organize their world and their interactions with others (Rosch & Marvin, 1975; Rosch et al., 1976). The study of categories and categorization schemes offers in this regard an important avenue for confronting the challenge raised by the pervasive involvement of data and data-management techniques in organizations and fields.

The tension between field-specific processes and decontextualized techniques of control is of course a pervasive theme in organization studies and, particularly, the branch of organization studies known as institutional theory. Field-specific processes often develop endogenously encoding experiential knowledge in a field, and other professional and organizational exigencies (Scott, 2007). By contrast, decontextualized techniques of control embody general principles of calculation and measurement that are largely field-independent (Fligstein, 1990; Kallinikos & Hasselbladh, 2009; Miller & O’Leary, 1987). While developing in particular social settings, such decontextualized techniques often diffuse across a variety of fields and organizational types (Abdelnour et al., 2017; Labatut et al., 2012), including public agencies (Bejerot & Hasselbladh, 2013) but also fields such as finance, travel or insurance (Shiller, 2003; MacKenzie, 2006). Our own study links to and, at the same time, renews the relevance of this type of research by bringing it to bear on organizational fields that are pervaded by data and the technologies and operations by which data are made to matter.

### **Technology and Organization**

We have so far showed how the apparatus of data-technologies-algorithms is linked to the establishment of objects and categories in organizations such as Last.fm. The data production process cycle we have unravelled (playcounts, artist objects, artist similarity networks, recommendations) is in most essential respects a process of categorization that is built on artist names as the basic objects and on

the higher-order categories of similar (and popular) artists. Playcounts link Last.fm with an atomized population of users while basic objects *qua* boundary objects shape the interaction patterns between users, developers and the platform. Such findings address the first two of the questions we raised at the frontend of this paper.

Yet, our third question concerning the wider organizational implications of these developments still requires further commentary. Organizations are broader social entities that embody much more than just technologically-embedded tasks and processes. Since its very beginnings as a scientific field, organization theory has construed organizations in terms of structural attributes such as jobs and positions, authority lines, resource compounds and knowledge systems and capabilities or in terms of processes that describe the entanglement of resource flows and actions, including routines, sustaining the production of goods and services. Over the years, such structural and processual attributes of organizations have variously been linked to the core technologies they deploy (Hickson et al., 1969; Mintzberg, 1979; Noble, 2017). This understanding of technology's involvement in organizations has implied a clear ontological separation between the technological and the social or organizational system as, for instance, in Mintzberg's (1979) distinction between the *operative core* and the administrative *superstructure* or in Hickson's et al.'s (1969) distinction of *operations technology* and *organization structure* (for a further and more recent discussion see Leonardi 2012; Leonardi & Barley, 2010).

The underlying idea has always been to find out how structural, processual and other attributes of organizations are shaped by the technological system and its features, and vice versa. While historically studied in the context of industrial technologies, the conception of technology and organization in these terms has by and large continued to underlie the study of information technologies and their organizational implications. One can retrace that separation even in Zuboff's (1988) work mentioned above and other studies on information technology and its link to social structure or processes (e.g.

Kallinikos, 1999, 2007; Labatut et al., 2012; Munir & Jones, 2004; Orlikowski, 1992; Zammuto, Griffith, Majchrzak, Dougherty & Faraj, 2007). Much current research on algorithms, AI and machine learning carries that tenor into our own age.

Organizations such as Last.fm and similar platform-like organizations are however different. In a sense, they collapse the difference between technology and organization. The standardization of the human interface and the automation of the backstage operations the apparatus brings about increasingly envelop the layout of organizational tasks and the processes with which they are associated and blur the distinction between humans and machines. The diffusion of smart and light technologies through which most users are currently linked to platforms and the web more widely makes the distinction fuzzier. Taken together, these developments reframe membership rules and the separation of organizations from their environments that many organizations, no matter how laboriously, have sought to maintain. Users are not employees nor are they simply customers. Critically, technologies are used in ways that reclaim the instrumentation of key organizational operations that have traditionally been performed by experts and work groups. The processes of categorization we have analysed in this paper and the process cycle with which it is linked is a case in point.

Much of what we describe in this paper demonstrates that the process of categorization is considerably hardwired to technologies and technological operations. At the same time, categorization is centrally linked to the production of data as a new and pervasive resource and medium characteristic of this hyper-technological age. Users are profiled on the basis of the music preferences they express through their clicking behaviour. The translation of cultural taste by other modes of expression (buying behaviour, social participation, income, class belongingness) to click data already marks a significant change of social modes of expression and identification and establishes an entirely new channel (data and their outputs) that links organizations to their surroundings. Once transformed to data, preferences undergo an alchemy of operations and can be added, subtracted, aggregated and computed to

generate other descriptions (classifications) of the social such as similar artists or similar users. This is how similarity but also popularity and trending scores produced daily and amassed on these platforms work.

Cast in this light, the difference between one set of operations that could be described as social from that of the operative technological core (Mintzberg, 1979) is reduced and the two sides mingle and interlace with one another in ways that make their separation a challenging task. Paraphrasing McLuhan (the medium is the message), one could go as far as to claim that the *technology is the organization* (McLuhan, 1967). This is certainly a hyperbole but a useful one that casts the current processes of organizing to a new light. It definitely calls for further research on these matters that can help unpack the empirical variability through which the fusion rather than simply the entanglement or imbrication of technology and organizations takes place (Leonardi 2012; Leonardi, Nardi & Kallinikos, 2012; Von Krogh, 2018; Zammuto et al., 2007).

## **Concluding remarks**

In this paper, we have described how categorization links to organization. We have focused on what we call the apparatus of data-technologies-algorithms and the ways it delivers the building blocks of organizing in contexts such as Last.fm, marked by the pervasiveness of data and operating as platforms. We have gone to some detail showing how the work of that apparatus reframes cultural and expert-based modes of categorizing by relying on the production of standardized types of data that are mashed up and assembled in several stages to provide the categories of the world on the basis of which final services are extracted and delivered.

While highly diffused throughout our societies, platforms remain badly understood as organizations. Our take on platforms differs substantially from how platforms have been studied in other fields such as economics (two-sided markets) or engineering. It provides a perspective for studying platforms



from an organizational point of view that focuses on the socio-cognitive work of categorization and the building blocks of organizing with which categorization has been linked. There is of course considerable variation out there. Last.fm is just a case. It is also possible to conjecture that the processes we describe here may have a different hold in other long-entrenched organizational and professional settings (Ekbia & Evans, 2009; Jones et al., 2018). While cultural and expert-based processes of categorization may persist, the developments we link with the diffusion of data and new formal and decontextualized techniques of data ordering warrant close examination that would allow us to identify changes in organizing and, ultimately, organizations.

## References

- Abdelnour, S., Hasselbladh, H. & Kallinikos, J. (2017). Agency and institutions in organization studies, *Organization Studies*, 38:12, 1775-1992.
- Aggarwal, C. C. (2016). *Recommender systems*. London: Springer.
- Alaimo, C. & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms, *The Information Society*, 33:4, 175-191
- Alaimo, C., & Kallinikos, J. (2019). Social media and the infrastructuring of sociality. *Research in the Sociology of Organizations*, 62, 289-306.
- Alaimo, C. & Kallinikos, J. (2020). "Recommender systems", in Beyes, T., Holt, R. and Pias, C. (eds.) *The Oxford Handbook of Media, Technology and Organization Studies*, Oxford: Oxford University Press
- Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.
- Bechmann, A., & Bowker, G. C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1).
- Bailey, D. E., Leonardi, P. M., & Barley, S. R. (2012). The lure of the virtual. *Organization Science*, 23(5), 1485-1504.
- Bejerot, E., & Hasselbladh, H. (2013). Forms of intervention in public sector organizations: Generic traits in public sector reforms. *Organization Studies*, 34(9), 1357-1380.
- Berger, P. & Luckmann, T., (1967). *The social construction of reality: A treatise on the sociology of education*. London: Penguin.
- Beverungen, A., Beyes, T., & Conrad, L. (2019). The organizational powers of (digital) media. *Organization*, 26(5), 621-635.
- Beyes, T., Holt, R., & Pias, C. (2020). *The Oxford handbook of media, technology, and organization studies*. Oxford: Oxford University Press.
- Boast, R., Bravo, M., & Srinivasan, R. (2007). Return to Babel: Emergent diversity, digital resources, and local knowledge. *The Information Society*, 23(5), 395-403.
- Borgmann, A. (1999). *Holding on to reality: The nature of information at the turn of the millennium*. Chicago: University of Chicago Press.
- Bowker, G. C. & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge MA: MIT Press.
- Ciborra, C. U. (2009). Imbrication of representations: Risk and digital technologies. In *Journal of Management Studies*, 43 (6), 1339-1356.
- Couldry, N. & Mejias, U. (2019). *The costs of connection*. Stanford: Stanford University Press.
- Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*. Cambridge, MA: Harvard University Press.
- Douglas, M. (1986), *How institutions think*. London: Routledge.

- Eco, U., (2000). *Kant and the platypus: Essays on language and cognition*. Houghton Mifflin: Harcourt.
- Ekbia, H. R. (2009). Digital artifacts as quasi-objects: Qualification, mediation, and materiality. *Journal of the American Society for Information Science and Technology*, 60(12), 2554-2566.
- Ekbia, H. R., & Evans, T. P. (2009). Regimes of information: Land use, management, and policy. *The Information Society*, 25(5), 328-343.
- Eisenhardt, K. M., (1989). Building theories from case study research. *Academy of Management Review*, 14(4): 532-550.
- Espeland, W., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, 113(1), 1-40.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62-70.
- Fligstein, N. (1990). *The transformation of corporate control*. Cambridge, MA: Harvard University Press.
- Flyverbom, M. (2019). *The Digital Prism*. Cambridge University Press.
- Flyverbom, M., & Murray, J. (2018). Datastructuring—Organizing and curating digital traces into action. *Big Data & Society*, 5(2), 2053951718799114.
- Foucault, M. (1970). *The order of things*. London: Tavistock.
- Gawer, A. (Ed.). (2009). *Platforms, markets and innovation*. Cheltenham: Edward Elgar Publishing.
- Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. Cambridge, MA: MIT Press.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Hacking, I. (2001). Aristotelian categories and cognitive domains. *Synthese*, 126(3), 473-515.
- Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. *New England Journal of Medicine*, 363(4), 301-304.
- Hannan, M. T., Pólos, L., & Carroll, G. R. (2007). *Logics of organization theory: Audiences, codes, and ecologies*. Princeton University Press.
- Hickson, D. J., Pugh, D. S., & Pheysey, D. C. (1969). Operations technology and organization structure: An empirical reappraisal. *Administrative Science Quarterly*, 378-397.
- Jones, M., Blackwell, A., Prince, K, Meakins, S., Simpson, A. & Vuylsteke, A. (2019) Data as process: From objective resource to contingent performance, in *Institutions and Organizations: A Process View*, Oxford: Oxford University Press.
- Kallinikos, J. (1999) Computer-based technology and the constitution of work: A study on the cognitive foundations of work, *Accounting, Management & Information Technology* (9:4), pp. 261-291.
- Kallinikos, J. (2007). *The consequences of information: Institutional implications of technological change*. Cheltenham: Elgar.
- Kallinikos, J., Hasselbladh, H., & Marton, A. (2013). Governing social practice. *Theory and Society*, 42(4), 395-421.
- Kallinikos, J., & Hasselbladh, H. (2009). Work, control and computation: Rethinking the legacy of neo-institutionalism, *Research in the Sociology of Organizations* 27: 257-282.
- Knorr-Cetina, K. (1999). *Epistemic cultures*. Cambridge, MA: Harvard University Press.
- Kornberger, M., Pflueger, D., & Mouritsen, J. (2017). Evaluative infrastructures: Accounting for platform organization. *Accounting, Organizations and Society*, 60, 79-95.
- Labatut, J., Aggeri, F., & Girard, N. (2012). Discipline and change: How technologies and organizational routines interact in new practice creation. *Organization Studies*, 33(1), 39-69.
- Lakoff, G. (1987). *Women, fire and dangerous things*. Chicago, IL: University of Chicago Press.
- Lena, J.C. and Peterson, R.A. (2008). Classification as culture: Types and trajectories of music genres. *American Sociological Review*, 73(5), 697-718.
- Leonardi, P. M., (2012). Materiality, sociomateriality and sociotechnical systems, in Leonardi, P. M., Nardi, B. A., & Kallinikos, J. (Eds.). *Materiality and organizing: Social interaction in a technological world*. Oxford: Oxford University Press, 25-66.
- Leonardi, P. M., & Barley, S. R. (2010). What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing. *The Academy of Management Annals*, 4(1), 1-51.
- Leonardi, P. M., Nardi, B. A., & Kallinikos, J. (Eds.). (2012). *Materiality and organizing: Social interaction in a technological world*. Oxford: Oxford University Press.
- MacKenzie, D.A., (2006), *An engine, not a camera: How financial models shape markets*. Cambridge, MA: MIT Press

- MacKenzie, D. (2017). A material political economy: Automated Trading Desk and price prediction in high-frequency trading. *Soc Stud Sci*, 47(2), 172-194.
- McLuhan, M. (1967). *Understanding media: The extensions of man*. MIT press.
- Miller, P., & O'Leary, T. (1987). Accounting and the construction of the governable person. *Accounting, Organizations and Society*, 12(3), 235-265.
- Mintzberg, H. (1979). *The structuring of organization: A synthesis of the research*. Englewood Cliffs: Prentice-Hall.
- Monteiro, E. & Parmiggiani, E. (2019). Synthetic Knowing: The Politics of the Internet of Things. *MIS Quarterly*, 43(1), 167-184.
- Munir, K. A., & Jones, M. (2004). Discontinuity and after: the social dynamics of technology evolution and dominance. *Organization Studies*, 25(4), 561-581.
- Negro, G., Hannan, M. T., & Rao, H. (2011). Category reinterpretation and defection: Modernism and tradition in Italian winemaking. *Organization Science*, 22(6), 1449-1463.
- Noble, D. (2017). *Forces of production: A social history of industrial automation*. London: Routledge.
- Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, 3(3), 398-427.
- Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435-1448.
- Perrow, C. (1986). *Complex organizations: A critical perspective*. New York: McGrawHill.
- Rogers, R. (2013). *Digital methods*. Cambridge: MA. MIT press.
- Rosa, J. A., Porac, J. F., Runser-Spanjol, J., & Saxon, M. S. (1999). Sociocognitive dynamics in a product market. *Journal of Marketing*, 63, 64-77.
- Rosch, E., (1975). Cognitive reference points. *Cognitive Psychology*, 7(4). 532-547.
- Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4). 573-605.
- Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3): 382-439.
- Rosenblat, A. (2018). *Uberland: how algorithms are rewriting the rules of work*. Oakland: University of California Press.
- Rosenfeld, L. & Morville, P. (2002). *Information architecture for the world wide web*. O'Reilly Media, Inc.
- Scott, W. R. (2008). Lords of the dance: Professionals as institutional agents. *Organization Studies*, 29(2), 219-238.
- Scott, W. R., Ruef, M., Mendel, P. J., & Caronna, C. A. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care*. Chicago: University of Chicago Press.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.
- Shiller, R. J. (2003). *The new financial order: Risk in the 21st century*. Princeton: Princeton University Press.
- Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3), 377-391.
- Star, L. S. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, 35(5), 601-617.
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1), 111-134.
- Timmermans, S. & Berg, M. (1997). Standardization in action: achieving local universality through medical protocols. *Social Studies of Science*. 27, 2: 273-305.
- Timmermans, S., & Epstein, S. (2010). A world of standards but not a standard world: Toward a sociology of standards and standardization. *Annual Review of Sociology*, 36, 69-89.
- von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4), 404-409.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 1-19.
- Yin, R. K. (2009). *Case study research: Design and methods*. London: Sage.
- Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization Science*, 18(5), 749-762.
- Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.

Zuckerman, E. W. (1999) The categorical imperative: Securities analysts and the legitimacy discount. *American Journal of Sociology*, 104: 1398–438.

Zwick, D., & Dholakia, N. (2004). Consumer subjectivity in the age of Internet: The radical concept of marketing control through customer relationship management. *Information and Organization*, 14(3), 211-236.

## **Author Biographies**

Cristina Alaimo is Assistant Professor (Research) in Digital Economy and Society in LUISS Guido Carli University, Rome, Italy. She holds a Ph.D. in Management, Information Systems and Innovation from LSE - London School of Economics and Political Science, UK. Her research focuses on the innovation brought about by data-based services and their consequences for organizations and society. Cristina's work has been published in journals such as *The Information Society*, *Journal of Information Technology* and *Research in the Sociology of Organizations*. Cristina is currently Visiting Research Fellow at Surrey Business School, University of Surrey, UK.

Jannis Kallinikos is Professor of Organization Theory in the Department of Business and Management at LUISS Guido Carli University, Rome, Italy where he holds the Cisco Chair on Digital Transformation and Data-Driven Innovation and Professor in the Department of Management at the London School of Economics and Political Science, UK. His research focuses on the impact of information and communication technologies on organizations and economic institutions. He has published widely in Management, Information Systems and Sociology journals and written several monographs including *The Consequences of Information: Institutional Implications of Technological Change*, Edward Elgar, 2007, *Governing Through Technology: Information Artefacts and Social Practice*, Palgrave, 2011. He has together with Paul Leonardi and Bonnie Nardi co-edited *Materiality and Organizing: Social Interaction in a Technological World*, Oxford University Press, 2012.