

NYU Press

---

Chapter Title: Data

Chapter Author(s): Melissa Gregg and Dawn Nafus

Book Title: Keywords for Media Studies

Book Editor(s): Laurie Ouellette and Jonathan Gray

Published by: NYU Press

Stable URL: <http://www.jstor.com/stable/j.ctt1gk08zz.20>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



NYU Press is collaborating with JSTOR to digitize, preserve and extend access to *Keywords for Media Studies*

JSTOR

Melissa Gregg and Dawn Nafus

Data play a major role in orchestrating contemporary power relations through the collecting capacities of knowledge-generating machines. For media studies, “data” is an increasingly important term as information gathered and shared through personal and public communication channels becomes subject to new kinds of tracking, quantification, and analysis. Media technologies like the smartphone combine multiple functions of broadcast, storage, transmission, and capture, turning everyday experiences into information that can be measured, sold, or used for political claim making. But discussion of “data” didn’t start in media studies. The terminology is drawn from traditions of information science, sociology, and the natural sciences. In these disciplines, recorded observations combine to create frameworks for understanding social phenomena and the behavior of populations, whether birds, humans, or microbes. The systematization of data in these disciplines previously required a human agent to conduct the analysis. Today computational machines are just as likely to provide the source of empirical revelation, as software packages and backroom analytical engines perform commands that allow for large-scale composition and representation of data. This automated assessment of data, the large-scale amassing of insights that is sometimes referred to as “big data,” can have the effect of stripping important contextual cues and details from the activities being measured. For example, geospatial data are time-stamped, but the

meaning being measured is often transported over time through space—a route, say—which likely involves sociocultural rhythms and meanings merely evoked by the data themselves, or perhaps erased entirely. Or consider graph/network data, which pretend they are time-free but are actually a snapshot of a specific moment in history. Simple attribute data that seemingly tell an eternal truth (fingerprints, blood type, or light from a star) are sampled once in time but are presumed to refer to an incontrovertible essence. Data have different types and functions depending on context; their meaning rarely remains fixed.

For most people, data tend to connote an individual record of activity. Using new types of media technologies, recording our activities can provide a source for reflection, self-enlightenment, and motivation. Looking beyond the individual, however, data can act as a shared record of human endeavor. The following is a list of further qualities we ascribe to data in this sense.

*Data are collected insights.* They begin with an individual fact—a *datum*, the Latin singular—and attract further instances to lay the foundation for an argument. Historically, the word has conveyed different meanings, but it has always referred to the tension between truth and persuasion.

*Data always have a date.* Date and data share the same etymological root. Data are recorded in time, and provide a trace, even when the recordings of time itself are erased through subsequent calculation. Patterns in time—cadences, steady declines, recurrences, and spikes—point to phenomena not otherwise recorded in conjunction with the data values themselves. At a technical level, the time stamp is a critical key that allows heterogeneous data sets to be brought together in the same calculation or visualization.

*Data gain significance through association.* “Data” is a word used in the singular and plural for good reason.

A datum does not stand on its own, but requires other datums to mean. They come together to say something. But to do so, first they must be assembled. This work of crafting association is necessarily rhetorical, since it is never possible to capture all information adequately. For data to mean something, the network of associations at some point must be cut.

*Data's potential is to facilitate narrative.* But whether they ultimately do so is a secondary question. An isolated activity that produces no evidence does not become data. It remains solitary, rogue, discountable, exceptional. It is a tree falling in the woods. The work involved in articulating data to elaborate new stories takes years, sometimes centuries. Data incite narrative bundles: culturally scripted accounts explaining what a sign indexes. Data have the capacity to destabilize the categories that underpin narratives, reinscribe them, or do both at the same time.

*Data may or may not point toward culturally stable referents.* Data point to a possible whole phenomenon, which in turn might not be a coherent assemblage. Take the example of an electricity monitor. It partially indicates energy consumption, but it might make only more evident the mysteries of what is actually involved in “energy consumption” beyond electricity. The impulse that data should be “actionable” is grounded in the assumption that data already have a social and cultural coherence—a routine of acting and knowing we are already trained to perform, and which data prompt us to do. Often they do not.

*Data are neither qualitative nor quantitative.* Data can appear as lines of text, or images, or categorical information, “qualitative” information that can potentially be counted, and rendered quantitative. Numbers, conversely, contain both symbolic and aesthetic qualities. When visualized, data “prepare the senses” in Michael Pryke’s (2010) phrase—their shapes contain qualities

that people respond to in ways beyond the merely intellectual. In Pryke’s study of financial data in stock market trading, visualizations worked as a kind of sensory prosthetic, bundled into the embodied gestures of anticipation in the practice of market trading.

*Media technologies capture data.* They provide the recording vehicles for activity, and means of communicating the stories told with data. In media studies, these stories tend to take three forms. The first contains *self-assembled information about individuals where data capture is self-nominated* and people have some say in crafting the narrative. In the Quantified Self subculture, for example, people choose to adopt tracking technologies such as wearable fitness monitors to record physical activity, heart rate, and sleeping patterns. Productivity software works similarly to record the screen activity of technology users so that an archive of practices can be generated for subsequent reflection. In both of these cases, data visualization and statistical measures are outputs that operate as points of reflection. This is their rhetorical power: data prepare us for an exploration about the body, the mind, or the senses that works in between our observation. Data allow us to alter aspects of a hidden lifeworld not always available to the conscious mind or witnessing eye. Compelling data prompt reform, improvement, reflection, or an aesthetic impulse.

*Second, media also capture information about individuals, assembled by others.* These are data that are captured and aggregated en masse by third parties for particular purposes, with or without an individual’s consent. When National Security Agency analyst Edward Snowden acted as a whistle-blower to reveal the extent of data surveillance conducted by the US government, individuals responded by claiming new rights to privacy to oppose such widespread monitoring of intimate life. Payment transactions, traffic routes, energy consumption, and phone conversations are some of the

most well-known data sets amassed by external bodies and institutions. This emerging context for popular governance is complicated by the fact that citizens are not always told about or actually understand the ways their data are collected. The most common justifications for the capture of nonidentifying behavior are the convenience of predictive services (e.g., Google Maps) or matters of civic patriotism, safety, and self-care (e.g., against terrorism or in response to natural disasters). Following Hurricane Sandy in 2012, Occupy Sandy activists collected on-the-ground information about the location and troubles affecting victims of the storm so that civic services could be mobilized. Without these door-to-door surveys, existing relief agencies would not have known who or how to help (Superstorm Research Lab 2013). The politics of data thus reflect broader tensions and inadequacies in the equitable provisioning of services in public life.

A third category of data collection and deployment is *community-oriented and purposive in nature*. Ventures in citizen science involve collecting and harvesting data to represent issues that may be difficult to learn about otherwise. Data activism allows minority groups a vehicle for telling urgent stories that are not in commercial interests to tell, from the impact of refinery pollution to the damage wreaked by fracking sites. Participatory geographic information systems aim to represent ownership or other types of claims for marginalized peoples. For motivated parties, data can be the evidence needed to secure more just and accountable social ends for the conduct of industry and government.

Beyond these three categories of data collection, the expansion of ambient means of gathering information, especially in contexts of professional work and management, harbors a range of in-between states of measure that are not fully voluntary or involuntary. In workplace programs of activity tracking, whether through

wearable tracking devices for fitness and health care benefits or email- or screen-based monitoring for enterprise security, employees are subject to varying levels of coercion and persuasion. This reflects a broader trend toward normalizing oversight of behavior by a technically mediated managerial gaze that captures both labor and lifestyle activities. It is unclear how much choice consumers or workers will have in this passive experience of surveillance as more of our everyday infrastructure becomes capable of monitoring activity through intelligent, embedded sensors.

In all of these instances, *data are collected with the intention to produce actionable insights—knowledge that prompts a response, even if that response is further reflection*. The difference between the forms of data collection lies in our awareness of or involvement in the process. When data are self-assembled, we may experience a feeling of control. But this feeling is haunted by an awareness that while data offer the capacity to tell stories about activities, we will never fully grasp the meanings of the data assembled. The notion of freedom frequently celebrated in the design of recording technologies is one that privileges individual will as the best kind of agency. Circumstances that force us to collect our own data—diabetes, allergies, sleep disturbances of mysterious cause—might involve a more begrudging obligation than freedom, but there remains scope for autonomy over when, and how, and, to some extent, what recording is performed. Conversely, when data are collected without assent, we become subjects of others' discourse. Our activities are recorded in the terms set by others, for the purview and measure of an external entity or authority. As Mark Andrejevic (2014) notes, a “big data divide” is emerging, where individuals have little chance to influence the terms upon which their information is gathered and used. This power asymmetry is one of the main points of concern for contemporary media studies.

The idea of an asymmetrical data economy in media studies (see also Nissenbaum 2009) reflects an inheritance from the early work of Michel Foucault (1978) and the notion of *pouvoir-savoir*, or power/knowledge. Over a series of studies, Foucault explained how modern institutions rendered activities visible and knowable by capturing them in discourse—by witnessing or describing behaviors in such a way that they could be named and categorized. Using media technologies today, this principle applies to the extent that accumulating data demonstrates new activities befitting new kinds of categories and narratives, and to the extent that these activities prompt external authorities to abstract the significance of individual narratives to generate new modes of regulation and order. But what has changed with the growing processing power of billions of connected data-capturing devices is that there is no human agent capable of adequately assessing the amount of identifying information people create in aggregate. Media studies needs new frameworks to understand a global marketplace and an international territory for governance that is distributed in complex ways, yet determined by large digital data sets.

Our research develops different metaphors and frameworks to answer this challenge. To explain the often unspectacular experiences of data exchange in everyday life, we are attracted to organic concepts, for example, the notion of *data sweat* (Gregg 2015a). Sweat is a natural phenomenon that happens to all of us. It describes an emission of meaningful information depending on context—weather, anxiety level, proximity to others, social engagement. It is also a form of information flow that toys with our ideas of control and agency, since sweat responds to different social and cultural cues. What media studies can sometimes miss, partly because of its focus on text and format, content and audience, is the difference that place makes in perception. More

recent theorists (e.g., Bachmann and Beyes 2013; Sloterdijk 2011) are beginning to identify the significance of environments in our engagements with media.

The current fascination with data metrics and analytics can be read optimistically as an interest in technology's role as facilitator for new kinds of stories. This is the last gasp of what has been called "participatory media." When communication technologies and people are equally mobile, we are no longer observing discrete bodies interacting with static media entities of transmission and aggregation so much as we are elaborating a hybrid relationship of occasional collaboration. The media studies to come will need to explain our involvement with data and their capturing devices as an accommodation, a cohabitation, a shared breath, a mutual dwelling.