Small data theorizing encompasses the ways communities of collaborators deliberate over the status of data through a set of three mediational logics that pose alternatives to datafication. Small data theorizing is a way of generalizing a framework for understanding the role of the collaborators in digital archiving and database design processes, which often involve both communication researchers and ordinary institutions, such as universities and museums. This framework for mediational logics extends the critiques of big data theory, and its monolithic claims about autonomous technologies and power in society, by elaborating the ways people actively maintain their connectedness to data in archival design. At the same time, it joins critical communication theory’s central concerns about power with a praxis. This can help all communication researchers in seeing how communities are theorizing practical alternatives to archives that are designed for big data and their ecosystems.

Keywords: Big Data, Critical Communication Theory, Digital Archives, Collaborative Design, Small Data

doi:10.1093/ct/qtaa029

Small data theorizing

There can be little doubting the importance of digital data to the contemporary global economy, governance, and the mediatization of everyday life for millions of people globally. On this phenomenon, critical communication theory has been at the forefront of understanding the rise of a political economy invested in the production and exchange of digital data at its grandest scale in order to serve the most nefarious interests of the largest governments and corporations globally. In their articulations, big data references not only the quantity of data, but also the qualities of a system that defines ontologies, epistemologies and phenomenologies of digital data according to a logic of datafication (Van Dijck, 2014). Datafication refers to the ways big data become the means and ends of digital archiving, an automated process that sacrifices the human ability to interpret the social world (Szulc, 2019). Whereas the quantity of big data is not threatening in itself, many critical communication theorists have sounded alarms in the past five years that datafication...
practices threaten free democracies, good governance, both private and public economic markets, consumer choices, social welfare and equal societies (Couldry & Hepp, 2016; Halperin, 2015; Langlois, Redden, & Elmer, 2015; Van Dijck 2014; Van Dijck, Poell, & de Waal, 2018). “What is at stake here is not simply our ‘trust’ in specific government agencies or single corporations, but the credibility of an entire ecosystem—an ecosystem that is fueled by a steady flow of billions of e-mails, video, text, sound, and metadata” (Van Dijck, 2014, p. 204). In this pioneering articulation of a critical communication theory agenda, digital data and datafication seem inseparable from an entire ecosystem.

Meanwhile, there have been communities of online users exploring alternative ecosystems by co-creating original digital archives. In this article, I outline three logics through which communities’ design practices are alternatives to the logic of datafication. Drawing on the various case studies in which communities and researchers are co-designing the technological infrastructure for an archival database and interface, their logics consider the role of community members in postulating the ways culture can be considered data, for whom, and by what means. Using Humphrey’s distinction between computational practice and social practice, these mediational logics are the foundational bricks upon for building a practical critical communication theory of data and digital archives through collaborative practices.

Consider, researchers aside, the number of communities of interest that pool their stories and artifacts into digital repositories and databases. Unlike a physical archive, the design of a digital archive—its database and interface—precedes and requires the transformation of the stories and artifacts into data (McGann, 2007). Communities design digital archives to further various objectives: from citizen science and civic media, to heritage preservation and social justice. They can involve a wide range of institutions, such as museums, local councils, universities and libraries, and organizational partnerships, such as professional nonprofits and subcultural fan groups. These are “ordinary” institutions in the sense that they frequently mediate between citizens and larger governance institutions in liberal societies (Kennedy, 2016, p. 11). What connects these geographically dispersed and socially diverse groups of communities and institutions are the challenges of designing original digital archives that allow their stories and artifacts to be collected, databased, and distributed through platform interfaces. “Collaborative iteration” describes the generalized process through which communities “take advantage of the native capacities of multiple platforms” in order to create new platforms (Clark, cited in Clark et al., 2014, p. 1452), but to do so requires deliberation and reflection on key theoretical questions about what data are, how they relate to social knowledge or human experience, and the value of digital data for community members. In this article, I consider them lay theorists of small data, with the modifier “small” used here to distinguish the human scale required to interpret data versus the computer scale to process big data. Small data in my formulation is relative and contextual in meaning, rather than an absolute quantity of data.
Articulating the mediational logics through which community members theorize small data extends critical communication theory about digital data specifically, and affirms an underlying investment in “practical theory” that unifies an otherwise-fragmented communication discipline (Craig, 2018, p. 290). In terms of critical communication theory, this article responds to a challenge launched nearly a decade ago to critically unpack assumptions that equated the power of datafication to the means and ends of digital data archiving (Boyd & Crawford, 2012). That theory of datafication, and the ways it assumes data ontologies, epistemologies and phenomenologies, follows a trajectory in critical communication theory to reveal structural relationships of power and domination (Fuchs & Qiu, 2018). Since then, there have been various studies of participatory design of digital databases, platforms, and analytics, all of which counter a monolithic theory of datafication, while also reuniting datafication theory with praxis. For the communication theorist Robert T. Craig (2018, pp. 295–296), the exchange between the social processes of cultivating communication and the “deliberative reflection on practical conduct” are the foundations of a practical theory that underpin all communication subfields, including “naturalistic communication science, critical studies, and applied science.” While small data theorizing intervenes in critical communication theory, then, its ethos more generally is in keeping with a practical theory for the discipline.

In order to unpack the mediational logics that guide small data theorizing in digital archive design, I first establish the critical communication framework by which archives express the social power of their designers. This framework has informed a number of critical communication theorists to argue that digital archives are both the cause and effect of the automated logics of datafication. Their work on the ontology, epistemology, and phenomenology of big data demonstrate the ways the social is already determined by the design of the technological infrastructure. From there, I posit small data theorizing as a way to bring the social back into theories of digital archives. This section summarizes a through-line in historical accounts of communities doing the collective work of archival design. This collective work, which emphasizes a multi-method approach to human-computer interactions, can be considered the basis of an alternative ecosystem for digital data in the future. The generalized logics by which communities theorize the alternative ecosystem comprise the next sections of the article. My generalizations are based on approximately a half decade of studies in which communities are designing digital archives through collaborative iteration practices. I personally came to understand these practices over a three-year period as a participant in a community-based design thinking process. I refer at times to these practices anecdotally to give a concrete example of theorizing, not to universalize the design experience or its socio-technical outcomes. While I acknowledge some of the limits of considering small data theorizing as a universal or global phenomenon, I hope that its articulation of a heuristic
for participatory community logics is productive alternative to what Kennedy and Bates (2017) warn is a monolithic notion of data power and encourage communication scholars to further theorize the political frameworks needed to support these alternatives in the future.

Critical communication theories of the archive and data

Data describe “facts on the lowest level of abstraction,” which allow people to make distinctions, compare, model, theorize, organize, or simulate their distribution (Krippendorf et al., 2016). The taken-for-grantedness of data’s facticity assumes they exist independently of individuals. Yet in critical communication theory, data cannot be separated from human agency or social power. Digital archives, in this same vein, are expressions of their designers’ orientation towards data as a reflection of the social world. As such, critical communication theory alerts us to the ways archives, with their databases and interfaces, reproduce social power through their design. These insights have been key to subsequent theorizations of the prevailing logics by which data are databased and deployed through digital platforms, generalized first in a Foreign Policy article as “datafication” (Cukier and Mayer-Schoenberger, 2013) and later developed in critical communication scholarship (Couldry & Hepp, 2016; Halperin, 2015; Langlois, Redden, & Elmer, 2015; Van Dijck 2014; Van Dijck, Poell, & de Waal, 2018). These insights have also underlined, however, an emerging set of practices to design digital archives that do not obey the logics of the status quo.

Long before the computer, archives communicated stories about a culture via the objects it held and organized, largely for expert users. Gitelman (2006, p. ix) writes, “Records and documents are kernels of humanistic thought, of the specifically modern hermeneutical project that has been associated since the 19th century with university departments of history and literature as well as many broader, less academic institutions of public memory, like libraries and museums, and other resonant forms of authoritative cultural self-identification, such as anthologies, reference books, bibliographies, and similar compendiums.” In other words, archives comprise both databases and interfaces. In the Western tradition mediated social power both through the collection of the objects as data representing a culture and through the interfaces created by the institutions that managed data uses. Communities attached to institutions—such as the museum, the library, and the university—defined the protocols for what objects would be included their archives, their ontology in a catalogue, and the ethics for accessing data. Communities vested in an archive thus shared in its social power as the members of a society authorized to communicate the stories embedded in data therein (Derrida, 1996). Hypothetically, the tight relationships between vested communities of experts and data contained in an archive would be democratized in a digital format that is neither limited by a physical space nor by an epistemic regime of the institutional sponsor. The term “database” itself references the possibility of overlapping data
ontologies that supersede the professional archivist’s authority over a singular catalogue (Harris, 2016). Similarly, “interfaces” are not stable objects, but infer a multiplicity of processes (Galloway, 2012).

Nevertheless, critical scholars argue, digital archives still mediate social power. In an early critique of the democratized database in digital archives, McGann (2007, p. 1588) notes the power of an interface to constrain the interpretation of database contents: “No database can function without a user interface, and in the case of cultural materials, the interface is an especially crucial element of these kinds of digital instruments. Interface embeds, implicitly and explicitly, many kinds of hierarchical and narrativized organizations.” Unlike a physical archive in which users can negotiate their content access and organization with professional gatekeepers, such as archivists, a digital archive controls these in absolute terms through its protocols. Critical communication scholars have theorized the multiple ways that technological infrastructures for digital archives determine the meaning of cultural contents via standardized systems of organizing data’s ontology (c.f. Galloway, 2004; Gillespie, 2012).

Following these protocols’ potentials, critical communication theorists have described big data as both cause and effect of a tectonic shift in epistemology and phenomenology, “a new paradigm for science and society” (Van Dijck, 2014, p. 198) that stemmed from faith in data’s communicative objectivity (Halperin, 2015). In the new paradigm, big data have agency to create knowledge and govern the social according to a logic of datafication that is oriented only to collecting more data that can pinpoint people into computational categories (Szulc, 2019). No longer inanimate objects held in the archive for expert access, big data render users into objects, which operate as “data doubles” (Cheney-Lippold, 2017) that “platforms concomitantly measure, manipulate, and monetize” in the absence of real citizens or consumers (Van Dijck, 2014, p. 200). Passive data flows can be harnessed to substitute ways of knowing populations without their conscious inputs or interventions (Crawford, Milner, & Gray, 2014). “Digital media power works in the epistemological dimension to reconfigure reality,” writes Packer (2013, p. 297), as computers come to replace humans in cognitive activities of data collection, assessment, guidance, direction, release, and prediction.

The gap between knowing and using data becomes fractionally ever smaller as automated algorithms take the place of the original programmer in making decisions, “an autonomous sense-making procedure” (Andrejevic, 2013, p. 14). Users’ autonomy are similarly pre-empted in data-responsive public spaces where sentient tools are configured to limit human experiences (Couldry & Hepp, 2016, pp. 129–137). Driven by a datafication logic, these “deep mediatizations” of digital data direct our taken-for-granted experiences of the social world to be “oriented to goals, driven by wider economic forces, that are different in type from the goals that embodied actors are able to have, unless, that is, they give up on their autonomy entirely” (2016, p. 142, emphasis added). The restriction, if not elimination, of human agency and autonomy for critical communication theorists of big data portend...
the totalization of what James Beniger (1989) predicted as a “control society” in which automation would replace human labor.

While big data in these theories seems self-serving as the means and ends of datafication, the political consequences of digital archives designed through datafication are grim. Digital data have become the raw material for capitalist appropriation in the 21st century; their economic use-value enriches the owners of archives that serve expanding commercial and governmental goals (Andrejevic, 2013; Srnicek, 2016). Big data reflect a political economy captured by “new concentrations of power” (Crawford, Milner, & Gray, 2014, p. 1667), identified specifically as “a few mega platforms driven by datafication logic” (Szulc, 2019, p. 15, see also Van Dijck, Poell, & de Waal, 2018; Srnicek, 2016) or as the “private corporate sector” more generally (Couldry & Hepp, 2016, p. 140, see also Striphas, 2015).

Datafication-driven archives operate through autonomous algorithms to reproduce this power through governmental structures that increase surveillance and minority profiling (Benjamin, 2019), and reduce democratic transparency and civic participation (Langlois, Redden, & Elmer, 2015), exacerbating social inequalities (Eubanks, 2018) and governing everyday social practices (Bucher, 2018; Van Dijck & Poell, 2013). “Far from being the mere owners of information, companies are becoming the owners of the infrastructures of society,” exclaims Srnicek (2016, p. 91). From this closed and all-encompassing political-economy portrayed in critical communication theory, there can be little room for alternative infrastructures.

Or can there? The questions of power and institutional structures that have defined critical communication theory since its interdisciplinary beginnings, guiding research that focuses “on changes in asymmetrical political and economic relations” (Melody & Mansell, 1983, p. 104). This prevailing agenda opened up questions about how dominant institutions develop and deploy new communication technologies, allowing “more and more types of information to be treated as marketable commodities” (1983, p. 114). At the same time, the focus on political and economic change has motivated critical scholars to search for viable alternatives to dominant institutional conditions under capitalism (Smythe & Van Dinh, 1983). Critical theory thus conjoins the critique of institutional structures with “transformative praxis that aims at social change towards a better world that defends and extends democracy and participation, and that works towards a good life for everyone” (Fuchs & Qiu, 2018, p. 227). It is this practical theory orientation that challenges digital communication technologies that are developed or deployed primarily to serve “capitalist accumulation or political administration” (2018, p. 223). By considering digital archives more broadly, theorists avoid reducing the social contexts of communication to the technology (Golumbia, 2013) or their “commercial mechanisms” (Kennedy, 2016, p. 28).

To summarize, critical theories of archives and archiving have revealed the ways that data collection and storage, and interface access and practices have always mediated the social power of institutions and their communities of experts. With digital archives, those communities have been marginalized by computational processes.
that, through the self-contained logic of datafication, collect more data to objectify humans and animate objects. Datafication pushed critical communication scholars to theorize digital data in light of an ecosystem in which the institutions that control the databases and interfaces of their archives. For alternative ways of conceptualizing digital data in an archive, I look to how grass-roots groups have theorized data as culture in designing their own digital archives.

**What small data theorizing is**

To depart from a monolithic understanding of data as big data demands an investigation of alternative ecosystems in which digital data are collected, stored, and accessed. These ecosystems predate the digital archive, giving some insights into the communicative contexts for considering culture as data. These insights have been plumbed more profoundly by communication researchers embedded in participatory design processes with lay communities. Whether focused on social media or infrastructural planning, these studies use multi-methods to interpret how these communities make sense of culture as data, a process I consider small data theorizing. As such, they provide a corpus for understanding digital data through mediational logics not contained solely by the logic of datafication.

While not abundant in the literature on archives, new histories of pre-digital archives (Flinn 2007; Gitelman, 2013; Gruber Garvey 2013) suggest a tight cultural relationship between lay communities and the cultural objects they wanted to consider data. Frequently established to create a counter-narrative to official cultural stories, lay communities looked to archives as a way to preserve an identity, a culture, or a movement (Flinn, 2007). In these cases, communities confronted a central design challenge: how to render culture into a data ontology that would be generally intelligible, while still allowing data to speak to the multiplicity of human experiences? For while poems, scrapbooked objects, and pictures could be abstracted as facts, they were also narratives that “cannot stand the tests to which they are subjected, i.e., the tests of verifiability, reliability, facticity, or representativeness” (Young, 2014, p. 191). By maintaining the connection between data as objective facts and data as analogous to situated and subjective experiences, these archiving communities were theorizing data.

Quite apart from a “media logic” through which media determine meaning, Humphreys (2018, p. 22) mobilizes the phrase “mediational logic” to mark “the shift in prioritization from the processes of datafication to mediation” in theorizing how data maintain their connectedness to human creators and users. This logic produces a “qualified self” that, unlike the data double produced by big data, communicates through social interactions. “Mediation, rather than datafication, reveals the ways that others feature prominently in the representations of ourselves,” she explains (2018, p. 23), describing a process that is necessarily dialectical in human interactions with digital technologies. Humphreys extends social theories of communication technologies in two ways. First, there is the interpersonal
communication around shared data that mediates culture. Second, there is the political economy that enables the technological infrastructure for a mediational logic. To this, Blackman (2015) offers a third way to distinguish digital data through a mediational logic of humans and technologies. Qualitative researchers offer lenses into “the micro-dynamics of data generation, and see how data traces can disclose patterns of activity missed if one is guided solely by metrics and computational forms of analysis based on predictive analytics” (2015, p. 248). She uses the phrase “small data” to describe the limited scale that permits qualitative researchers to interpret, analyze, or theorize data; hers is a direct response to Boyd and Crawford’s (2012) complaint that digital data have become synonymous with big data.

One place to find small data theorizing is in the design phase of digital archives, when the human agency for establishing protocols and standards for data processing is most capacious. Ethnomethodological studies conducted by Neff, Tanweer, Fiore-Gartland, and Osburn (2017) and Kennedy (2011), for example, revealed the ways data scientists and web designers continuously engaged in ethical judgments and reflexivity in co-creating interfaces. Bivens (2015) found further longitudinal evidence that coders struggled over the database categories and protocols as a form of resistance to the biased or discriminatory outcomes from computer automation. Struggles over digital archive design articulate the tensions between new groups of experts—data scientists and web designers—and automated systems that control the ways culture is encoded as data through a database and interface. These struggles over the ontological status of culture as data are often flattened in critical communication theories of big data.

Bringing these frameworks together may allow a deeper look into contemporary case studies in which researchers embed in participatory processes that consider culture as data. These studies fall within various paradigms—such as engaged scholarship, community-based design, and action research—but share insights into the social processes that challenge deep mediatization and datafication. Villanueva et al. (2017, p. 475) place such studies in communication under the umbrella of communication infrastructural theory, because they generate theory from within social settings that stress “the establishment of academic-advocacy partnerships in order to meaningfully bridge research and practice.” In their framework, “communities” are discursively formed among members of the group (2017, p. 477), allowing for flexible application and adaptation to digital settings. For example, online communities are marked by fluid membership, typically involving both “maintainers” and “visitors,” that contribute to digital archives only once or infrequently (Bighash, Oh, Fulk, & Monge, 2018). Communities may emerge through the process of designing a digital archive, rather than being a product of an pre-existing community (Baker & Huber, 2013). These designing communities are my focus for a further exploration of the communicative practices that are theorizing small data.
Three mediational logics in small data theorizing

The following sections aim to unpack the mediational logics that communication researchers have documented in designing digital archives with communities. Through these logics, I use the term “theorizing” to shift the locus for interpreting data to the communicative and social practices among groups. Indeed the ontological, epistemological, and phenomenological questions raised by critical communication theorists about big data have been raised by the participants in different design processes. Small data theorizing is thus an alternative framework for critically theorizing an archival ecosystem not subsumed by datafication.

The corpus for this deeper exploration comes from the case studies in approximately a decade of books and peer-reviewed publications, which themselves can be found in a digital archive managed by the Communication Institute for Online Scholarship, “a not-for-profit service created by the communication field itself, responsive to the public trust of science and scholarship.”

What becomes evident in looking at the studies is that there has been no standard methodology for doing this kind of engaged design research with communities. In part, this is because different communities themselves use a variety of “DIY practices” to design their archives (Baker & Huber, 2013). Also in part, communication researchers may focus on certain stages of a design process. For example, some researchers have focused their studies on co-creating technical infrastructures for organizing culture as data in a digital database (i.e., Srinivasan, 2017) while others have engaged in a collaborative questioning about the status of data to represent culture through an interface (i.e., Leurs, 2017). Still others have participated in gathering groups to co-design data mining algorithms or applications (i.e., Kennedy, 2016).

In considering this corpus, I am limiting my focus to the social practices involved in digital archival design, rather than the specific context, geography, or motivation of each study and the design practice each describes therein. The study of design practices are themselves useful to articulate for the growing number of communication researchers who engage in these practices as participants, including myself. My experiences are meant as illustrative of logics in the corpus rather than as a case study from which to draw generalizations.

What follows are three mediational logics that guide small data theorizing in designing digital archives. They are:

1. A logic of multiplying ontologies to order and classify data;
2. A logic of mindful consent to collect, store, and use data;
3. A logic of social use-values towards an alternative data economy.

These logics may overlap in their practical application. Yet each one points towards mediational logics that are alternatives to the datafication logic that seems to colonize critical theories of digital data. Alternatives are generative in critical communication theory. They can expand or challenge ways of seeing how power...
operates in societies. They provide new interpretative lenses for analyzing objects and their relationships to social knowledge and systems of power, while also creating new practical challenges for designers. Taken as a whole, the mediational logics of small data theorizing pose new ways of imagining an alternative ecosystem for digital data generated through social practices.

A logic of multiplying ontologies
What counts as data? How should they be ordered to reflect the community, its collective knowledge and differences? Communities confront these questions in designing informational architectures which, by definition, structure data into classificatory chains and paradigms of meanings. Although the chains and paradigms form ontologies which always “assume a certain context and perspective of some community” (Schoop, Moor, & Dietz, 2006), the logic of datafication affixes singular ontologies in opaque algorithms to reify social hierarchies (Gillespie, 2012).

In contrast, small data theorizing confronts the challenge of making data reflect plural and competing ordering systems. I experienced this theorizing practice through a collaborative sorting workshop in which 15 participants took stories, images, and other contents in a database and reorganized them into ontological trees based on consensus. In the process, groups had to decide how to translate social knowledge into data. The trees split to grow complex horizontal branches and vertical root systems that represented local themes and authorial genres, two entirely different ways of seeing culture as data. We resolved this tension by creating two ontologies, which overlapped in the search results, but also were displayed in the search interface as two organizing principles for data. Through a similar logic, community groups have found multiple ways to structure digital data in their socio-technical systems to reflect shared “knowledge, capabilities, and skills, as well as social and cultural relationships and connections” (Alexiou et al., 2016, p. 203).

Although some communities find technical means to assist everyday people in writing flexible structured query languages (NoSQL) that manage data (Srinivasan, 2017, pp. 42–43), the proposition that everyday people can classify data in meaningful paradigms is most often realized through popular tagging and metadata practices. These allow different users to participate in data taxonomies, finding new connections and pathways through data. In their sum, tags may create a vernacular folksonomy of horizontal data connections that resist normative knowledge frameworks. For example, much of the efforts to incorporate citizen scientists and participants into community-based research has necessitated a logic that allows for communal sense-making around the organization of data using their own local stories. In allowing students, faculty, and local residents to combine their metadata and linkages in a database of video fragments, news articles, and social media posts, De Moor (2014) found that the community identified new classificatory patterns which evaded the “often hierarchical, inward-looking, and monodisciplinary way of
doing science” (2014, p. 109). As an added bonus, the collaborative process kept these publics more engaged with their archive.

In turn, communities theorize small data’s ontologies within a reflexive process that precedes and postdates their development. Srinivasan (2017) charts this longer trajectory in his work with different indigenous tribes that wrestled, for example, with non-native categories for “art” or “the sacred.” He describes the processes of community-driven consensus that foreground how “key topics and themes in the life of a community” only emerge from repeated reflections “on the collective traditions, practices, values, priorities, and epistemologies of their community” (2017, p. 135). Theorizing presumes interpersonal communication, but also repeated negotiations as consensus may shift over time. In one tribe, he notes that consensus about the meaning of objects in their archive changed as the tribe itself became more open to outsiders and computers. Srinivasan notes that by conceptualizing data ontologies as fluid and in flux, communities overcome “the limitations of a ‘user’ to embrace the diverse knowledge systems, values, and protocols that are part of community life” (2017, p. 136).

A logic of multiplying data ontologies creates practical and theoretical challenges as communities sort through “a lot of categories” based on popular additions without subtraction (Gorjounova & Shulgin, 2006, p. 239). Human-generated tagging taxonomies may compete for placement in search results, generating new issues of hierarchy and authority over the ordering of social knowledge. In our own work, for example, the phrase “Mardi Gras Indian” competed with the phrase “maroon society” to describe a local cultural practice. Though both terms existed in the database, the former displaced the latter in search results simply because more participants recognized it. At the same time, a community that embraces multiplying ontologies may take steps towards deepening connections in their data, using the back-end architecture as its own whiteboard interface. Over time, some connections will grow thinner or break from disuse, while others will grow and adapt to new identifications. One such project found that data categories that “nobody wanted to be identified with disappeared” (2006, p. 249), once again reaffirming the importance of recurring human intervention in creating the ontologies.

These fluid and many ontologies may frustrate communities of experts who have typically been in charge of naming and organizing archival data in order to be legible to future generations. Uricchio (2009) asks what is to be done when there are no longer authors speaking to the meaning of artifacts and when the connective links between the artifacts are lost, because: “Their rootedness in community and collaborative interactions and their responsiveness to an ever-shifting present [which] give them an unique quality as Zeitdokumenten, finely grained embodiments of culture” (2009, p. 141). To this, and the potential big data solutions to Uricchio’s complaint, small data theorizing approaches preservation in ways that may satisfy the community of now, but not always that of the past or the future.
A logic of mindful consent

How do communities maintain connectedness to data? How should participants treat each other in collecting, storing, and using data? A logic of mindful consent summarizes the variety of ways community practice reflexivity on the ethics of digital data collection and access. Whereas big data disconnects data from humans and their communities (Metcalf & Crawford, 2016), and thus subsumes the social via algorithms, small data theorizing prioritizes the social, focusing instead on the ways people opt-in to be data producers and users.

Hui (2015, p. 238) posits that communities develop a “technics of care” in guiding the normative practices in their digital archives because their data represent an “exteriorization of our memories, speeches, and movements.” These technics of care are the result of small data theorizing about how to upload or download contents to which people feel personally attached but want to share. Theories about authorship and appropriation may guide various technical pieces of the interface, for example: opt-in language and “terms of consent” for all users, mandatory documentation of image or audio releases, and regulated levels of access to parts of the archive. These considerations became part of the interface for our media library, for example, by requiring that users document permissions and ownership of artifacts as types of intellectual property. Theories about ownership and usage generate shared ground-rules for collaborative design and a pedagogy for future consent.

The culture of “good faith collaboration” in Wikipedia, for example, guides how participants manage conflicts over data (Reagle, 2010). Those who consent to share their knowledge must also be mindful of disagreements over their contributions. The community of participants guides participants how to negotiate, compromise, and dissent with each other within an ethical stance that differences over the meaning of data must be tolerated, but that there are many routes to consensus in resolving those. The continual reaffirmation of a deliberative production process that stresses mindful consent over any particular type of data in the archive encourages civility and tolerance for different ideas about data. The permissions process for media artifacts, in our case, allowed various options for sharing and reuse on the site, as well as notifications allowing an owner to share the work when recombined with others. Notifications give the owner an opportunity to renegotiate the terms of use. The changing contents of the database thus cannot be separated from the logic through which all recognize that data are subject to practices of group consensus.

A logic for mindful consent underlines users’ activities in a wide range of socio-technical practices with small data. By building in multiple places in which users opt-in to automated systems of data production and uses, communities may give citizens agency over data platforms (Kennedy, Poell, & van Dijck, 2015). Automation in data practices can proceed in a context in which participants know how data is meaningful to others, while also placing some degree of responsibility on designers to make participants understand the terms of their consent. Couldry, Fotopoulou, and Dickens (2017) theorize that when communities regularly...
participate in designing data algorithms for their own needs, they produce “real social analytics” that better reflect their own changing social realities. Real social analytics allow “users to retain control over their conditions of existence, through reflexive adjustment to the parameters of their online categorization and measurement” (2017, p. 126). At the same time, ethical practices of recurring consent and consensus building can also set the parameters for individual users’ control.

Operating through a logic of individual consent that is also mindful of the whole, community-based initiatives confront challenges over which data to preserve and how. As “alternative data practices” situate data in human knowledge and experiences, small data is not self-evidently meant to be collected, stored, or shared (Leurs, 2017). Consent is mediated by affective relationships to data artifacts (Baker & Huber, 2013), and those can take precedence over an impulse to archive everything or open access to everyone. As such, Flinn (2007, p. 169) argues:

> [E]ven if we agree that these materials tell the story of not just one community but tell the stories of everyone, it does not follow that we should presume access to those materials as a right. With collections that belong to a community, deposit and public access are not a right but of matter for negotiation, partnership and encouragement.

The results of community negotiations over data can be exclusive as much as inclusive. In one telling illustration of this, members of the Zuni indigenous community, working with outside designers and researchers, theorized the need to exclude data from a digital archive shared with a network of outside museums (Srinivasan, 2017). While project members struggled to ensure data did not become isolated from the oral stories and traditions that gave them meaning in the community, the group also decided to exclude data that elders considered too “religiously sensitive,” much to the chagrin of museum professionals who saw the data as part of the tribe’s heritage. This case exemplifies how popular participation in data practices may limit as much as broaden the scope of small data, as well as the competing notions of authority over small data among various stakeholders.

**A logic of social use-values**

What are the social forces guiding the production and exchange of data? How do communities support an alternative data ecosystem? In critical theories of big data, all human activity is subject to objectification, whether extracted from data workers or condensed in the form of data doubles, that can be exchanged as a commodity (Andrejevic, 2013). In contrast, small data theorizing prioritizes social use-values over market exchange-values. Our own process continually stressed cultural preservation, recognition, pluralism, and personal voice as social-use values in the project, even if these were in tension and subject to more consensus-building. The logics of multiplying ontologies and mindful consent contribute to an ecosystem in which data remain subjectively connected to user communities. Social use-values may frame the structuring of digital archives’ sustainability within ordinary institutions.
Fluid and many ontologies and repeated practices of mindful consent create archival databases and interfaces never totally detached from the discursive community of lay designers. These design choices encourage users to deliberate over what count as data, and their epistemic relation to social knowledge, but also how to reiterate the archival design. In my own experiences, collaborative iteration inferred a dynamic but unending process of communicating how to maintain social relationships with data. For example, participants in a digital archive created by a collaborative of citizen journalists and college students focused on the social-use value of local stories in iterating a shared platform. Yet when new contributor groups struggled to see how their non-narrative image or audio is a story, they pushed participants to re theorize the use-value of stories over other values, such art or information. As new stakeholders emerged and retreat from the design process, they re theorize how data reflect shared needs. These theories can be materialized through new architectural forms or revamped protocols for the archive.

Needs are not necessarily exchangeable. De Waal (2017) characterizes the process of community-based digital design as a “platform” in its original sense: a stage for playing out the social through group formation and relations. For him, collective needs and goals are negotiated through give-and-take conversations that occur online and offline as a shared “dramaturgy” in which participants take different roles in having agency over the platform (2017, p. 128). In his work documenting how community stakeholders would reimagine an urban parcel of land, the digital archive reflected the changing relationships between the groups as they “revalued” the property as a communal living space. The methods of production engaged in co-designing and co-creating, as well as co-managing the archive itself, reveal “intangible assets” in data that are not transferrable (Alexiou et al., 2016, p.203).

For example, Long, Baker, Istvandity, and Collins (2017) described how a community of popular music fans formed around their “affective” connections to data that represented a subcultural identity. They designed a digital archive that they hoped would have a use-value in preserving that identity, despite internal differences among group members. Through these dialectical dynamics of digital data creation and reflexive community practices, it is not a tautology to assert that the product of community-based design practices is the community itself.

Through the mediational logic of social-use values, the gift of participants’ time and effort on the collective archive project maintain this ecosystem. Repeated and ongoing design processes as a practical matter involve dedicated periods of immaterial labor among community members. “Despite the distinctiveness of the community, it is the people who build the platform and make it work, whether by contributing with products or by discussing, evaluating, or participating otherwise” (Goriunova & Shulgin, 2006, p. 259). All of the logics of small data theorizing require human labor, but efforts can be distributed in a variety of ways, from “a time-consuming participation model, to one with occasional bursts of activity” (2006). Participation as a social use-value thus may impact other mediational logics of
platform design, particularly when and how mindful consent is required in the technical architecture.

At the same time, this ongoing gift labor in the context of communities coming into and out of design processes leaves the question of the role of so many institutions that may wish to support or participate in these digital archives, and relatedly, what Kennedy (2016, p. 190) calls an effort “to do good with data.” The fact that so many communication researchers are embedded in the design practices they are studying would infer that academia plays an outsized role in creating alternative archive infrastructures, but to this researchers have added entities representing local government, neighborhood planning, and heritage and preservation (Couldry et al., 2017; De Waal, 2017; Flinn, 2007, respectively). These institutions can seek to align with amateur design groups to further their own mission or outreach efforts. Memoranda for understanding ensuring open data, contributor releases and retractions are three mechanisms that these institutions can uphold in sustaining digital archives when popular participation wanes or designers quit. As Kennedy (2016, p. 217) argues, these practices are not necessarily toppling the structures for big data power, they illustrate that alternative data structures are indeed possible.

Small data theorizing in a critical communication praxis

Articulated through a big data logic of datafication, digital archives would seem to pose an existential threat to ways of representing and knowing the social world. By following the ways the technical infrastructures for data archives mediate power relations, many critical communication have raised the specter of a big data ecosystem that squelches alternatives. In this sense, small data theorizing opens alternatives in conceptualizing the digital ecosystem. Communication researchers, working often as co-participants with communities engaged in designing their own digital archives, have studied how social practices reanimate data’s cultural meanings. The collaborative practices used to create a database or an interface demonstrate a reflexive theorizing, as Leurs (2017, p. 140) posits: “that digital data cannot be expected to speak for themselves, that data do not emerge from a vacuum, and that isolated data on their own should not be the end goal of a critical and reflexive research endeavor.” Small data theorizing preserves data’s facticity, while restoring their orientation towards human agency and experience.

In this article, I have organized small data theorizing according to three mediational logics for digital archive design practices. Such a framework both extends a critical understanding of digital archives as means for alternative understandings of data, but also reunites critical communication studies’ commitment to a practical discipline. The reinsertion of the social considerations around data as culturally meaningful can guide other practitioners in developing technical infrastructures that embrace the multiplying ontologies, ethics of mindful consent, and social use-values found in these case studies. An exploration of the real applications of mediational logics also shows the tensions and contradictions that people face when
trying to build an archive that represents diverse and dynamic communities over time. Limited gift labor and the goals of partnering institutions may constrain the sustainability of digital archives developed through small data theorizing. These considerations should not deter communication scholars from working with communities to theorize data. As Craig (2018, p. 295) argues, “Theory can be nonetheless useful insofar as it draws attention to matters that should be considered in deliberative reflection on practical conduct.” The challenges faced by the lay theorists of small data have laid the groundwork for future deliberations over what a digital archive might be for communities that differ in their discursive formation.

None of the mediational logics for interpreting small data should be seen as a panacea to the dark implications of big data archiving practices and their datafication logic. This is not a call to rally behind every popular design practice as morally good or resistant to data power, despite the monolithic values critical theorists ascribe to big data. In my own experience, the use of Google keywords and Google analytics are already standard additions to community-generated platforms, thereby capturing the resistant potential of our practices in the big data ecology. Similarly, many of the design participants in the case studies seem to be the first to critique their own limited scope for negotiation the meaning of data and how to grant mindful consent within the constraints of unpredictable gift labor. At the same time, the identification of mediational logics across a number of studies must add nuance to early claims positing their absolute hegemony. Institutional arrangements also exert their own pressures on how digital archives survive and thrive in articulating the social use-values and ethics of their founding communities. In some ways I have found that the energies dedicated to inventing an archive—its database or interface—have to be rechanneled when activated among new stakeholders. In this case, practices for regular collaborative iteration has ensured that digital archives remain connected to social knowledge and participants’ experiences.

Communication scholars also might apply the mediational logics of small data theorizing to envision how a digital archive might be sustained or governed at scale. Researchers working with communities to develop technical protocols that embrace these logics have sometimes extended the politics of designing groups to advocate for widespread open-data regulations or peer-production networking standards (Lovink & Rasch, 2013; Srinivasan & Fish, 2017). While these structural considerations are beyond the scope of this article, the effort to expand the mediational logics of small data theorizing to constrain, if not eliminate a datafication logic, is vital for the future.

Acknowledgements

I am grateful to my colleagues from the Phyllis Taylor Center for Design Thinking and community partners who invited me to be a part of the iterative and ongoing design process for the ViaNolaVie digital archive since 2015.
Funding

Funding for this project was received from the Louise K. Riggio and Carnegie Foundation endowed professorship in social innovation and entrepreneurship at Tulane University.

Notes

1 See http://cios.org/ (accessed January 2, 2019). Unfortunately, I have no insight into the logics for the design or operation of this archive.

2 A case study elaborating these design practices among Zuni participants can be found in Srinivasan (2017, pp. 168–177), though he has also researched these practices among Somali refugees and indigenous Mexican communities.

References

Alexiou, K., Agusita, E., Alevizou, G., Chapain, C., Zamenopoulos, T., & Turner, J. (2016). Asset mapping and civic creativity. In I. Hargreaves & J. Hartley (Eds.), The creative citizen unbound: How social media and DIY culture contribute to democracy, communities and the creative economy (pp. 181–204). Bristol, England: Policy Press.

Andrejevic, M. (2013). Infoglut: How too much information is changing the way we think and know. New York: Routledge.

Baker, S., & Huber, A. (2013). Notes towards a typology of the DIY institution: Identifying do-it-yourself places of popular music preservation. European Journal of Cultural Studies, 16(5), 513–530. doi:10.1177/1367549413491721

Beniger, J. (1989). The control revolution: Technological and economic origins of the information society. Cambridge, MA: Harvard University Press.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the New Jim Code. Cambridge, MA: Polity.

Bighash, L., Oh, P., Fulk, J., & Monge, P. (2018). The value of questions in organizing: re-conceptualizing contributions to online public information goods. Communication Theory, 28(1), 1–21. doi:10.1111/comt.12123

Bivens, R. (2015). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. New Media & Society, 19(6), 880–898. doi:10.1177/1461444815621527

Blackman, L. (2015). The haunted life of data. In G. Langlois, J. Redden, & G. Elmer (Eds.), Compromised data: From social media to big data (247–272). New York: Bloomsbury.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technical, and scholarly phenomenon. Information, Communication & Society, 15(5), 662–679. doi:10.1080/1369118X.2012.678878

Bucher, T. (2018). If. . . then: Algorithmic power and politics. London, England: Oxford University Press.

Cheney-Lippold, J. (2017). We are data: Algorithms and the making of our digital selves. New York: NYU Press.

Clark, J., Couldry, N., de Kosnik, A., Gillespie, T., Jenkins, H., Kelty, C., . . . van Dijck, J. (2014). Participations: Participatory promise of contemporary culture and politics. International Journal of Communication, 8, 1446–1473.
Couldry, N., Fotopoulou, A., & Dickens, L. (2017). Real social analytics: A contribution towards the phenomenology of the digital world. The British Journal of Sociology, 67(1), 118–137. doi:10.1111/1468-4446.12183

Couldry, N., & Hepp, A. (2016). The mediated construction of reality. London, England: Polity.

Craig, R. T. (2018). For a practical discipline. Journal of Communication, 68(2), 289–297. doi:10.1093/joc/jqx013

Crawford, K., Milner K., & Gray M. L. (2014). Critiquing big data: Politics, ethics, epistemology. International Journal of Communication, 8, 1663–1672.

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data and how it’s changing how we think about the world. Foreign Affairs, 92(3), 28–32.

Derrida, J. (1996). Archive fever: A Freudian impression. Chicago, IL: University of Chicago Press.

De Moor, A. (2014). Expanding the academic community research-building bridges into society with the Internet. In T. Denison, M. Sarrica, & L. Stillman (Eds.), Theories, practices and examples for community and social informatics (pp. 86–111). Melbourne, Australia: Monash University Publishing.

De Waal, M. (2017). Heritage as platform. In A. Asseldonk et al. (Eds.), Street values: The new playing field for preservation practice (pp. 116–133). Amsterdam, the Netherlands: Reinwardt Academy.

Flinn, A. (2007). Community histories, community archives: Some opportunities and challenges, Journal of the Society of Archivists, 28(2), 151–176. doi:10.1080/00379810701611936

Fuchs, C., & Qiu, J. (2018). Ferments in the field: Introductory reflections on the past, present, and future of communication studies. Journal of Communication, 68(2), 219–232. doi:10.1093/joc/jqy008

Galloway, A. (2004). Protocol: How control exists after decentralization. Cambridge, MA: MIT Press.

Galloway, A. (2012). The interface effect. London, England: Wiley.

Gillespie, T. (2012). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), Media technologies: Essays on communication, materiality, and society (pp. 167–194). Cambridge, MA: MIT Press.

Gitelman, L. (2006). Always already new: Media, history, and the data of culture. Cambridge, MA: The MIT Press.

Gitelman, L. (Ed.). (2013). ‘Raw data’ is an oxymoron. Cambridge, MA: MIT Press.

Golumb, D. (2013). Communication, critical. Communication and Critical/Cultural Studies, 10(2–3), 248–252. doi:10.1080/14791420.2013.806145

Goriunova, O., & Shulgin, A. (2006). From art on networks to art on platforms. In J., Krysa (Ed.), Curating immateriality: The work of the curator in the age of network systems. New York: Autonomedia.

Gruber Garvey, E. (2013). Facts and FACTS: Abolitionists’ database innovations. In L. Gitelman (Ed.), ‘Raw data’ is an oxymoron (pp. 89–102). Cambridge, MA: MIT Press.

Halperin, O. (2015). Beautiful data. Durham, NC: Duke University Press.

Harris, K. D. (2016). Archive. In B. Peters (Ed.), Digital keywords: A vocabulary of information society and culture (pp. 45–53). Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctvct0023.8
Hui, Y. (2015). A contribution to the political economy of personal archives. In G. Langlois, J. Redden, & G. Elmer (Eds.), Compromised data: From social media to big data (226–246). New York: Bloomsbury.

Humphreys, L. (2018). The qualified self: Social media and the accounting of everyday life. Cambridge, MA: MIT Press.

Kennedy, H. (2011). Net Work: Ethics and values in web design. Basingstoke, England: Palgrave.

Kennedy, H. (2016). Post, mine, repeat: Social media data mining becomes ordinary. Basingstoke, England: Palgrave.

Kennedy, H., Poell, T., & van Dijck, J. (2015). Data and agency. Big Data & Society, 2(2). doi:10.1177/2053951715621569

Kennedy, H., & Bates, J. (2017). Data power in material contexts: An introduction. TV & New Media, 18(8), 701–708. doi:10.1177/1527476417720034

Krippendorf, K. (2016). Data. In K. Bruhns Jensen, R. T. Craig, J. Pooley, & E. Rothenbuhler (Eds.), The international encyclopedia of communication theory and philosophy (pp. 1–6). London, England: Wiley. http://doi.org/10.1002/9781118766804.wbiect104/full

Langlois, G., Redden, J., & Elmer, G. (Eds.). (2015). Compromised data: From social media to big data. New York: Bloomsbury.

Leurs, K. (2017). Feminist data studies: Using digital methods for ethical, reflexive and situated socio-cultural research. Feminist Review, 115(1), 130–154. doi:10.1057/s41305-017-0043-1

Long, P., Baker, S., Istvandity, L., & Collins, J. (2017). A labour of love: The affective archives of popular music culture. Archives and Records, 38(1), 61–79. doi:10.1080/23257962.2017.1282347

Lovink, G., & Rasch, M. (2013). Unlike us reader: Social media monopolies and their alternatives. Amsterdam, the Netherlands: Institute of Network Cultures.

McGann, J. (2007). Database, interface, and archival fever. PMLA, 122(5), 1588–1592.

Melody, W. H., & Mansell, R. (1983). The debate over critical vs. administrative research: Circularity or challenge? Journal of Communication, 33(3), 103–116. doi:10.1111/j.1460-2466.1983.tb02412.x

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. Big Data & Society, 3(1), 1–14. doi:10.1177/2053951716650211

Neff G., Tanweer A., Fiore-Gartland, B., & Osburn, L. (2017). Critique and contribute: A practice-based framework for improving critical data studies and data science. Big Data & Society, 5(2), 85–97. doi:10.1089/big.2016.0050

Packer, J. (2013) Epistemology not ideology or why we need new Germans. Communication and Critical/Cultural Studies, 10(2), 295–300.

Reagle, J. (2010). Good faith collaboration: The culture of Wikipedia. Cambridge, MA: MIT Press.

Schoop, M., Moor, A. D., & Dietz, J. L. (2006). The pragmatic web: A manifesto. Communications of the ACM, 49(5), 75–76. doi:10.1145/1125944.1125979

Smythe, D., & Van Dinh, T. (1983). On critical and administrative research: A new critical analysis. Journal of Communication, 33(3), 103–116. doi:10.1111/j.1460-2466.1983.tb02412.x

Srinivasan, R. (2017). Whose global village? Rethinking how technology shapes our world. New York: NYU Press.
Srinivasan, R., & Fish, A. (2017). *After the Internet*. Cambridge, England: John Wiley & Sons.

Srnicke, N. (2016). *Platform capitalism*. London, England: Wiley.

Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies, 18*(4–5), 395–412. doi:10.1177/1367549415577392

Szulc, L. (2019) Profiles, identities, data: Making abundant and anchored selves in a platform society. *Communication Theory, 29*, 257–276. doi:10.1093/ct/qty031

Uricchio, W. (2009). Moving beyond the artefact: Lessons from participatory culture. In M. van den Boomen, S. Lammes, A.-S. Lehmann, J. Raessens, & M. T. Schäfer (Eds.), *Digital material: Tracing new media in everyday life and technology* (pp. 135–146). Amsterdam, the Netherlands: Amsterdam University Press.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society, 12*, 197–208. doi:10.24908/ss.v12i2.4776

Van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication, 1*(1). doi: 10.17645/mac.v1i1.70

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The platform society: Public values in a connective world*. London, England: Oxford University Press.

Villanueva, G., Gonzalez, C., Son, M., Moreno, E., Liu, W., & Ball-Rokeach, S. (2017). Bringing local voices into community revitalization: Engaged communication research in urban planning. *Journal of Applied Communication Research, 45*(5), 474–494. doi:10.1080/00909882.2017.1382711

Young, S. (2014). *Changing the wo(r)ld: Discourse, politics and the feminist movement*. London, England: Routledge.