

# The Coerciveness of the Primary Key: Infrastructure Problems in Human Services Work

CHRIS BOPP, University of Colorado Boulder, USA

LEHN M. BENJAMIN, Indiana University-Purdue University, Indianapolis, USA

AMY VOIDA, University of Colorado Boulder, USA

The premise of a primary key is simple enough: every record or row in a table should have some number or string that can uniquely identify it. Primary keys are essential for linking data spread across database tables, and for looking up and retrieving data from specific records. Yet for an identifier that seems so straightforward and uncontroversial, we find myriad ways that this unassuming bit of infrastructure has an outsized influence in human services work. Through case studies of the organizational networks of two nonprofit human services organizations, we find that different stakeholders use variants of identifiers to support work practices that are far more complex and social than the linking of tables or the lookup of data. Yet we also find that the low-level technical properties of the primary key are often coercive, forcing end-users to work on the infrastructure's terms—influencing the nature and order of the work, creating new forms of work, and influencing the tenor of the relationships among stakeholders. The technical abstractions of the underlying infrastructure, then, start to become the de facto public policy. We conclude by offering design provocations for better supporting identification across a variety of contexts.

Additional Key Words and Phrases: Human Services Organization; Identifier; Infrastructure; Interjected Abstraction; Nonprofit Sector; Primary Key

## ACM Reference Format:

Chris Bopp, Lehn M. Benjamin, and Amy Volda. 2019. The Coerciveness of the Primary Key: Infrastructure Problems in Human Services Work. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 51 (November 2019), 26 pages. <https://doi.org/10.1145/3359153>

## 1 INTRODUCTION

The primary key is one of the first concepts covered in introductory database design texts. The premise is simple enough: every record or row in a table should have some number or string that can uniquely identify it. Primary keys are essential for linking data spread across database tables and for looking up and retrieving data from specific records. Sometimes primary keys are relatively unique and anonymous numeric strings (e.g., a client ID number) and sometimes primary keys are less anonymous (e.g., a U.S. Social Security Number) or even less unique (e.g., a first and last name). Yet for a special instance of an identifier—a data point that establishes a connection to an object or individual—that seems so straightforward and uncontroversial, we find myriad ways that this unassuming bit of infrastructure has an outsized influence in human services work. What are the implications, for example, when an individual client is assigned two different client ID numbers by

---

Authors' addresses: Chris Bopp, University of Colorado Boulder, ATLAS Institute, 1125 18th St. 320 UCB, Boulder, CO, 80309, USA, [chris.bopp@colorado.edu](mailto:chris.bopp@colorado.edu); Lehn M. Benjamin, Indiana University-Purdue University, Indianapolis, IU Lilly Family School of Philanthropy, 301 University Boulevard, Indianapolis, IN, 46202, USA, [benjamin@iupui.edu](mailto:benjamin@iupui.edu); Amy Volda, University of Colorado Boulder, Information Science, 1511 University Ave. 478 UCB, Boulder, CO, 80309, USA, [amy.volda@colorado.edu](mailto:amy.volda@colorado.edu).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2019/11-ART51 \$15.00

<https://doi.org/10.1145/3359153>

separate, incommensurate databases, each one required by a different funder of the human services organization? What are the implications when an individual client provides different pseudonyms to different human services organizations that want to aggregate her/his data to improve service provision within a community?

Through case studies of the organizational networks of two nonprofit human services organizations—including 37 interviews with 43 staff members from an HIV/AIDS service organization’s network and a homelessness service organization’s network—we find that different stakeholders use variants of identifiers and primary keys to support work practices that are far more complex and social than the linking of tables or the lookup of data. Yet we also find that the low-level, technical properties of the primary key—including that the primary key cannot be changed, cannot be null or empty, and must be unique within its context—are bubbling up through the infrastructure and forcing end-users to work on the infrastructure’s terms, a class of infrastructure problem that Edwards et al. [24] call *interjected abstractions*.

In this research, we take a broad-based, work practice approach to the study of a specific facet of information infrastructure. Instead of bounding our unit of analysis around a given technology (e.g., the relational database and its primary key), we bound our unit of analysis around what we call *database practices* and *primary key practices*. This more expansive unit of analysis enables us to consider work practices that should inform system design but that do not currently rely on the technologies in question. We mirror, in particular, Volda et al.’s study of nonprofit sector database use that interrogates not just databases that are ‘technically’ databases, but that also include spreadsheets and paper forms used as databases [71]. We similarly expand our analysis of the primary key to include the myriad identifiers in whatever kind of information system they might appear that are used to support primary key work practices such as organizing, sorting, linking, and accessing records. Given this approach, we use the terms ‘database’ and ‘primary key’ with a limited degree of rhetorical elasticity in the work presented here—sometimes stretching their original technical definitions, but within the bounds of related practice. In taking up this unit of analysis and its related rhetoric, we are able to more holistically and empirically interrogate the tensions between the work for which stakeholders are appropriating primary keys and other identifiers and the influence of the technical infrastructure on that work.

## 2 LITERATURE REVIEW

### 2.1 Infrastructure & Infrastructure Problems in Human-Centered Computing

Information infrastructures are as old as the stones used by the British Empire to count their subjects [67]. While the power dynamics between those who define what *counts* for being counted and those who are *themselves* counted has continued to be a pervasive attribute of information infrastructures (e.g., [37, 45, 56]), the material forms of those infrastructures have continued to evolve [21]. From stones [67] to census tracts [56] to spreadsheets [21, 68] and databases [21, 36], the materialities of information infrastructures shape work practice [22, 46]. And particularly so in a contemporary society that values data-driven thinking and decision-making.

Researchers have focused extensively on characterizing the sociotechnical nature of information infrastructures, particularly its classification systems (e.g., [8, 11, 12, 43, 52]) and systems of measurement (e.g., [6, 49, 73]). As Star asserts, if you overlook the social and political nature of information infrastructures, you miss "essential aspects of aesthetics, justice, and change" [61]. For example, while the process of filling out a death certificate and declaring a cause of death may seem to be factually straightforward, the information infrastructure around this task is imbued with religious and ethical values as well as influenced by doctors’ work practices [12]. Additionally, Volda et al. [73] argue that the selection of particular units of measurement is highly political and prioritizes

the needs of certain stakeholders over others, which is especially difficult given the asymmetric power relationships in the human services context. Systems of classification and measurement exert clear political influence, then, over human experience. The experience of marginalization through systems of classification and measurement is pervasive enough that Bowker and Star have given a name to the experience, calling it torque: "the twisting that occurs when a formal classification system is mismatched with an individual's biographical trajectory, memberships, or location" [12]. Indeed, as Mol has argued: "The point of asking what is being counted is not to argue that counting is doomed to do injustice to the complexity of life. This is certain. The point, instead, is to discover how and in what ways" [42].

These systems of classification and measurement, when implemented in database systems, are formalized by those databases' schemas. The database—as infrastructure—then, imposes an additional requirement on the schema, that one field be declared the primary key to do the additional work of organizing, sorting, linking and accessing records. Although not often the focus of sociotechnical research in information infrastructures or classification systems, analytic interrogations of the database's primary key are one way to gain new insights into the functioning and implications of database systems. Raley notes the key role played by the primary key in linking records about an individual across databases: "...provided a set of different databases are networked and provided that they share the same means of establishing individual identification, so that a single unit (an individual or number) can be identified consistently across a range of data sets with a primary key" [51]. And Ribes and Jackson [53] further foreshadow the critical role of the primary key in linking data back to the real-world construct(s) that those data are intended to represent—in their case, the primary key provides the crucial link between a water sample and a specific stream. While the stream's primary key is simple and mundane, if the identifier were to become detached from the rest of the data, the entire system would break down because scientists would be unable to associate the water sample with its source.

Information infrastructures are particularly essential objects of study for the fields of computer-supported cooperative work and human-centered computing because they are performative; that is, the infrastructures, themselves, exert influence on the world around them. As Bowker asserts, "the database itself will ultimately shape the world in its image: it will be performative" [11]. This performativity has been studied in many contexts including environmental monitoring [53], flora and fauna [11], medicine [49], and emergency food systems [73]. As Pine and Liboiron have argued, the influence of data tools and practices are crucial for human-centered computing researchers and designers to take into consideration: "Since computing technologies such as databases, algorithms, and information entry interfaces, are designed around measurement; the development of measurements and the politics they embody can shape HCI design before it has even begun" [49]. And as Taylor has argued, systems design and implementation is also a process of 'world making' [65]. Design decisions matter not just for the affordances of the final artifacts, but in the assumptions they perpetuate about what is possible in the world and for shaping the kinds of futures that people imagine, advocate for, and work towards [65]. Within the nonprofit domain, Volda et al. [73] argue that the database systems used by organizations and tuned to collecting particular data based on external stakeholder needs become performative by shaping "the future of the organizations and institutions that we rely on to promote the public good and remedy social injustice." And critically, the impact of information systems on public policy may be felt unevenly by stakeholders who are powerful enough to define categories and schemas, and those—often clients—who are not [12, 62].

Because infrastructure underlies many systems layers in the design of computing, what Star and Ruhleder refer to as its "embeddedness" [62], infrastructure has an outsized influence on system design and user experience; it is "sunk" into, inside of, other structures, social arrangements and

technologies" [62]. Edwards et al. [24] characterize three "infrastructure problems" that derive from this design influence. First, infrastructure can create problems of *constrained possibilities*, in which "design choices taken by the infrastructure may preclude entirely certain desirable user experience outcomes." Second, infrastructure can create problems of *interjected abstractions*, in which "technical abstractions in the interface may appear in the conceptual model exposed to users." Third, infrastructure can create problems of *unmediated interaction*, in which "users may have to interact directly with the infrastructure to accomplish their goals." These infrastructure problems, largely unpacked by Edwards et al. [24] as challenges for user interface design and user experience, harken back to considerations of the materiality of information infrastructures [21]—that the embeddedness [62] of the technical abstractions is, itself, a reflection of the materiality of information. It is this theoretical bridging among multiple uses and audiences of the term "infrastructure" that we return to in our discussion, as it provides a productive vantage point for interrogating the infrastructural abstractions that force human services staff to work in their image.

This research also responds to Bietz and Lee's call for a richer understanding of how collaborative work is infrastructured by information ecologies comprised of multiple databases [7]:

The tendency to dismiss situations where organizations depend on imperfectly interoperable databases as merely inefficient legacy systems is likely glossing over insights about just how multiple databases support not only different types of work but also different perspectives and priorities [7].

Here, we unpack how the micro-scale player of the primary key has an outsized influence on human services work practices across the proliferation of databases [71] of the intra- and inter-organizational work of human services provision. In doing so, we highlight the importance of focusing analytic attention—not just on those infrastructures that might otherwise be rendered analytically invisible—but on the technical assumptions underlying the material instantiations of those infrastructures.

## 2.2 The Data Work of Human Services Provision

Human services organizations are organizations "whose principal function is to protect, maintain, or enhance the personal well-being of individuals by defining, shaping or altering their personal attributes" [31]. Human services represent the largest portion (73%) of United States federal spending—providing critical services such as workforce development, healthcare, and child welfare [20]. But rather than provide many of these human services directly, over the last fifty years, federal, state and local governments have increasingly contracted with nonprofit organizations to provide these services [57]. For example, the Urban Institute estimates that in 2012 governments paid close to \$81 billion to human services nonprofits for service provision [48]. With this funding has come increased expectations for accountability and evidence of impact, which inevitably means collecting data, which we have referred to elsewhere as *data compliance work* [5]. This pressure from funders runs alongside nonprofit professionals' own need for data that will help them identify problems, adjust practices, and coordinate with other providers to better meet the needs of individual clients.

But collecting and using this data to meet these external and internal demands is not as straightforward as it seems, in part because the data needs of funders and the data needs of human services staff are often different [3, 4, 23, 33]. Even when the same data could be used to meet both sets of needs, the data infrastructure requirements of funders make it almost impossible to do so [5, 9, 23, 33]. Consequently, most human services organizations focus on meeting funder data demands often at great cost and often without a clear benefit to the people they serve [3, 15, 23, 58]. All of this data work rests on top of limited technical capacity and resource constraints that

make the vision of data-driven human services provision outside the reach of most organizations [5, 9, 13, 14, 33, 39–41, 69, 71, 72].

In the past several years, data demands have gotten even more complicated for human services organizations, as they are now asked to aggregate data across organizations to track progress in addressing problems at the community-level [5, 9, 25, 38, 66]. Here, new data issues have emerged, requiring what we have referred to elsewhere as *data coordination work* [5]. Schoech [55] has predicted that the general trend toward big data will require human services organizations to develop greater interoperability of data systems and that this will require "predetermined data definitions, standards, protocols" and "user authentication and identity management tools to ensure data security, client privacy and confidentiality."

Privacy concerns add additional layers of complexity to the information infrastructures in some organizations that are legally mandated to ensure confidentiality. Even in other settings, where confidentiality is not legally required, privacy concerns strain the relationships that organizations have with clients as clients raise concerns about their rights in this new data sharing environment [60]. Although it is difficult to generalize about human services organizations given their diversity, on the whole, they tend to be more professionalized and bureaucratic than other nonprofits and, consequently, clients have less control in the organization, including over their data [30]. Further, in human services work, data collection and service provision occur simultaneously, rendering clients' conditions—homelessness, mental illness, HIV positive status—more visible and open to inspection, visibility that is all the more acute when the client is experiencing a socially stigmatized condition. The asymmetric power relationship foregrounded by the provision (or not) of services is a powerful frame for the experience of data collection (see also [10]). Clients often respond to mitigate this new vulnerability by limiting what they disclose [60]. Carnochan et al. [16] and De Witte et al. [19] found that collecting valid data, then, depends on working to build trusting relationships with clients—what we have elsewhere referred to as *data confidence work* [5]—so that clients are willing to fully disclose information about themselves. Finally, although data collection processes can and do vary across human services settings, in many human services organizations these data collection processes are extensive and so intertwined with service provision that they almost become the service, itself [5, 59].

Additional challenges of data work in human services provision also include the following: competition among organizations [34, 64], evolving priorities in public policy and associated funding [5], and the various ways that "doing good" can be operationalized [4, 70].

Overall, there is a significant need to understand the role and challenges of data-driven work in the nonprofit and human services context. As the pressure for these organizations to become more data-driven intensifies (e.g., [32]), significant questions have arisen about the effects of data-driven work practices on these organizations, their staff, and the clients they serve [3–5, 9, 15, 23, 33, 58]. Researchers have started to piece together a puzzle of various factors that are influencing or positioned to influence the nature of data-driven work in this context, including the influence of funders (e.g., [2, 9, 18, 23, 57, 63]), other organizations in the surrounding networks [35], and the public policy field [5]. This research contributes a missing piece to the puzzle in this line of scholarship: exploring how information infrastructure and its abstractions influence data-driven work.

### 3 METHODS

We conducted case studies of the organizational networks of two human services organizations—one in the HIV/AIDS services policy field and one in the homelessness services policy field—to better understand the information ecosystems of data work in the nonprofit sector. Elsewhere, we present findings related to the macro-level influences that the broader policy field has on information



infrastructures and work practices, including the ways that federal policies are implemented at the state and local level, the choice of funding tool used, and the assumptions about public policy problems and solutions [5]. Here, we turn to interrogate the role of a much more micro-level player, the database primary key, in the data work of human services organizations.

### 3.1 Case Selection

We selected two organizational networks in which we anticipated that the phenomenon of interest—the data work—would be rich and intense [47]. We used three criteria to select a focal organization for each network: (1) a nonprofit that received government funding to ensure that the information ecosystem would be undertaking *data compliance work*; (2) a nonprofit working in a policy field where there was an effort to achieve community-level outcomes to ensure that the information ecosystem would be undertaking *data coordination work*; and (3) a nonprofit serving a population that faced some societal stigma to ensure that the information ecosystem would be oriented toward *data confidence work*. We chose focal organizations working in two different social service areas (policy fields), so we could understand the degree of transferability of findings across service contexts.

One case centers around the information ecosystem and organizational network of an HIV/AIDS services organization. The focal organization in the HIV/AIDS case is a community organization in a county with a population of approximately one million people. Like many other community organizations founded in the late 1980s in response to the AIDS crisis, this organization has a diverse portfolio: service provision for those diagnosed with HIV/AIDS, including case management, mental health and addiction counseling, and medical services; free testing and risk-reduction counseling; and advocacy for the rights of persons living with or affected by HIV/AIDS. At the time of the study, the organization had nearly 50 staff members, had served approximately 1,200 HIV positive individuals, and had provided approximately 3,500 free HIV tests. The organization had received support from a range of foundations and donors, though the majority of its funding (62%) had come from county and state government.

A second case centers around the information ecosystem and organizational network of a homelessness services organization. The focal organization in the homelessness case, also founded in the late 1980s, is situated in a county with a population of about 400,000 people. This organization is one of three in the city that provides services to adults experiencing homelessness, including street outreach, overnight sheltering during the winter months, and transitional and permanent supportive housing year-round. At the time of this study, the organization also had approximately 50 staff members and 2016 census figures suggest that the organization contributed to serving a population of approximately 700 homeless individuals in the county, though these census figures are widely considered to under-represent the population [54]. The majority of this organization's revenue came from a combination of private sources (40%) and government grants (30%) through two city governments, the county government and, indirectly, from the federal government (as passed through a local coordinating organization).

### 3.2 Data Collection

We collected data through in-depth case studies within each focal organization, as well as with key informants at organizations in the focal organizations' surrounding information ecosystem. The analysis reported here is drawn from data collected through 37 interviews with 43 human services staff (Table 1). The semi-structured interviews broadly covered topics around data collection, entry, aggregation, use, and sharing.

In each focal organization, we conducted semi-structured interviews with a cross-section of staff who interact in a variety of ways with the organization's data, including frontline staff in

Table 1. Informants by Case

		<b>HIV/AIDS Services Case</b>	<b>Homelessness Services Case</b>
<b>Focal Organization</b>	Frontline staff, program managers, and data oversight	15 interviews (n=17): A1-A15	6 interviews (n=7): H1-H2, H8-H11
<b>Organizations in the Surrounding Network</b>	Staff from referral organizations	5 interviews (n=6) across 4 organizations: A16, A18, A20-A22	4 interviews (n=4) across 4 organizations: H4, H6-H7, H12
	Staff from data aggregating organizations (e.g., funders)	4 interviews (n=6) across 2 organizations: A17, A19, A23-A24	3 interviews (n=3) across 2 organizations: H3, H5, H13

different programs who are responsible for collecting data; program managers who are responsible for reporting this data up to senior staff; and quality assurance or other staff who provide oversight of data.

We also conducted semi-structured interviews that provided a more expansive view of the broader information ecosystem across each organizational network, including interviews with staff from other nonprofit organizations that were members of referral networks, with whom the focal organization shared clients and client data, as well as with staff from organizations who were responsible for receiving and aggregating data upstream (e.g., funders). In the HIV/AIDS case, this included staff from the county and state; in the homeless services case, this included staff at a nonprofit intermediary and the county.

This sample of informants provides a broad view of the organizational infrastructures and data work of the network of human services organizations in each case, though it does not represent the perspectives of all organizational stakeholders. For example, while we involved many frontline staff who work directly with clients on a day-to-day basis, we are unable to represent accounts of client experiences. Instead, our vantage point in this work enables us to understand how data work affects staff members' interactions with clients—the nature of their relationships with clients and their ability to render effective and supportive services to clients.

To better understand the complexities of the information infrastructures across each of these organizational networks as well as to clarify any ambiguities or discontinuities within or across informant accounts, we also employed two additional methods of data collection, used within the context of the semi-structured interviews:

- (1) To the extent allowable by the organization, we collected data artifacts that documented the information systems experienced by and characterized by informants, including data collection forms, the data dictionaries and database user manuals that described what staff should enter into a particular database, and the websites of the vendors who supplied the information systems and/or technical assistance to the organization.
- (2) In each focal organization, we also worked with the person most centrally responsible for the oversight of data to develop a data journey map for that organization and its network

[1]—flow diagrams of how data moved into, through, and out of their organization, including information about *what* data, *where* those data are organized and managed, *from whom* the data originated, and *to whom* the data are passed off.

We also observed coordinating meetings within each organizational ecosystem, at which service providers and funders discussed service coordination and data sharing. These data are not analyzed here but were used to help identify key stakeholders for additional interviews across each focal organization's network.

To support the distributed fieldwork of this interdisciplinary research team, the principal investigators of the larger research program (the second and third authors of this article), conducted a phase of joint pilot fieldwork at the first focal organization before the data collection effort reported here. This joint pilot fieldwork included an initial interview with a key informant, a facility tour, and two focus groups with a breadth of staff at the focal organization. We used this pilot fieldwork to develop an initial shared understanding of the setting, the work being carried out, and the infrastructures supporting this work; to begin to establish a common language around infrastructure and data work; as well as to identify common constructs of interest. Researchers then divided up the remaining data collection. Researchers met almost weekly throughout data collection to coordinate the ongoing, iterative refinement of the interview protocol and to coordinate a consistent sampling of informants across the organizational networks of the two cases.

### 3.3 Data Analysis

We analyzed data iteratively and moved between phases of inductive analysis grounded in the data and deductive analysis grounded in the research literature (see e.g., [17]). To begin, we read through the transcripts for each case separately, identifying instances of data work while attending to similarities and differences across the two cases. As with most studies, more than one interpretive frame can be used to understand the data, illuminating some aspects of a case while downplaying others [50]. In our first stage of analysis, we identified themes across the data suggesting the influence of the policy field on data work practices in human services organizations and their networks (published in [5]). In our second stage of analysis, we identified prominent, cross-cutting themes related to the role of identification, identifiability, and identifiers in those same instances of data work. We report on the details of the latter analytic process in what follows.

In our second stage of analysis, researchers inductively coded all interview transcripts for any broad relationships that the data work had to themes of identification, identifiability, and identifiers. Researchers met almost weekly to discuss the emergent coding categories and the relationships among them. Through these ongoing discussions, we began to see a distinction between: (1) coding that was descriptive of work practices that were implicated in the use of identifiers (e.g., that identifiers were important in counting clients) and (2) coding that was descriptive of the kinds of assumptions about the use of these identifiers in work practices (e.g., that critical identifiers needed to be collected first).

Our second round of coding, then, focused on distinguishing between these two broad categories of codes and developing additional granularity to our understanding of each of them. In this phase, researchers also generated memos using the guiding questions, "What work is being supported or constrained here?" and "What assumptions about identifiers are supporting or constraining these work practices?" Through the continued weekly discussions, particularly as supported by these memos, the role played by different variants of identifiers came to the fore, which served to highlight the workarounds of stakeholders and how the normative assumptions of the primary key—as a special and influential instance of identifier—influenced work practices. Our third round of coding focused on interrogating even more specifically what variants of the identifiers and



primary keys were implicated in these work practices and what infrastructural properties were influencing both the work practices and the contextual assumptions.

## 4 RESULTS

### 4.1 Variants of Identifiers Used as Primary Keys

*4.1.1 Variants Across a Design Space of Uniqueness and Anonymity.* In relational database design, the primary key is required to be *unique* to enable information retrieval through the identification of specific records of interest and to allow relationships to be represented across tables within a database through linking and cross referencing. Primary keys are also predominantly assumed to be *anonymous*, not directly identified with a person, as database design best practices commonly suggest using a numeric identifier automatically generated by the database, itself. In the human services context, however, these assumptions do not always hold. Organizations appropriate multiple variants of identifiers and primary keys—both in the degree of anonymity afforded as well as the degree of uniqueness of the identifier and/or the individual identified. Different variants have different affordances, which support different kinds of work.

In the human services context, primary keys typically refer to a row of data in a table that corresponds to a client or establishes a link between the client and an instance of service provision (e.g. meeting with a case manager or accessing a food pantry). Informants never used the specific, technical term "primary key"; that construct emerged from our analysis as having particularly useful explanatory power (see [28]) for explaining the assumptions and constraints around the use of identifiers. Instead, informants used language such as "unique identifier," "ID," or they described specific types of IDs like drivers' license numbers or U.S. Social Security Numbers. Informants used similar language regardless of whether they were referring to a primary key in a database or a similarly-functioning identifier in a spreadsheet or on a form, which mirrors our broader unit of analysis, studying identifiers used for primary key practices regardless of the material form of infrastructure supporting those practices.

Through our analysis, we found that identifiers used as primary keys in this human services context vary in the degree of *anonymity* they afford—the extent to which the associated client can be identified and linked back to the rest of the data. Example identifiers sometimes used as primary keys that vary along this dimension of anonymity include the following:

- **Anonymous Identifiers** such as randomly-generated or incremented numeric strings are maximally de-identified and generally considered to prohibit the reassociation of data with the individual represented by the data.
- **Obfuscated Identifiers** are created through hashes of legally-identifying information to render that data less identifiable (e.g., taking the last two letters from a first name and appending the last two letters of a last name and the day of the month the individual was born). Obfuscated identifiers enable some degree of anonymity while also theoretically enabling the identifier to be re-matched to an individual—or a small group of potential individuals.
- **Pseudonymous Identifiers** exist in fields defined to contain legally-identifying information, but may not contain legally-verifiable information. Pseudonymous identifiers may be used consistently over time, linking consistently to the same individual even though that individual might not be directly identifiable; pseudonymous identifiers may also differ over time, providing an inconsistent reference to an unidentifiable individual.
- **Legal Identifiers** such as names and U.S. Social Security Numbers are maximally identifiable and often require an additional level of verification against government-issued documents.

Through our analysis, we also found that identifiers vary in the degree to which they are *unique*, though in two distinct ways that are highly entangled in practice: (1) in the extent to which they are unique in the context of a given database, as is typically required in the design of relational databases, but also (2) in the extent to which they reference a unique individual. Example identifiers sometimes used as primary keys varying along this entangled dimension of uniqueness include the following:

- **Not Necessarily Unique Identifiers** are identifiers like names that can often be linked to a *unique* individual but are *not guaranteed to be unique* within a database (e.g., an instance in which two people with separate records in the same database share a common name).
- **Aliased Identifiers** are *unique* identifiers with respect to the database (e.g., an anonymous identifier might have been generated when a client presented to receive services on a first visit) but that are *not guaranteed to uniquely* represent a given individual (e.g., if a second anonymous identifier was generated for the same individual on a subsequent visit).
- **Deduplicated Identifiers** are *unique* identifiers within a given database that are also *guaranteed to uniquely* represent a given individual—that is, the unique identifier has also been deduplicated.

The two entangled manifestations of uniqueness become key to understanding some of the infrastructure problems that emerge in this context and so we'll return to this entanglement in our discussion.

All of the identifiers present in our data are situated at the intersection of both dimensions. That is, all identifiers have different affordances with respect to anonymity and uniqueness (e.g., anonymous and aliased identifiers, anonymous and deduplicated identifiers, legal and not necessarily unique identifiers, pseudonymous and not necessarily unique identifiers, etc.) The data does not include all possible permutations of identifiers—and it seems intuitively likely that not all permutations may be possible. But we stop short here of suggesting or creating prescriptive links between the two dimensions so as not to constrain the designerly imagination moving forward.

*4.1.2 Variants as Instantiated Across Different Classes of Information Systems.* The information infrastructures used by the organizations in this research are, like those of many other nonprofits, creative assemblages of many types of systems—referred to elsewhere as homebrew databases [71]. These homebrew databases include various enterprise-level or custom database systems; spreadsheet applications used as databases; and paper-based instantiations of databases, often forms that mirror the data-entry workflow of another database where the data will be entered later. Though the term "primary key" is typically only used by database administrators to refer to a special class of identifiers used in formal database systems, here we extend its use to describe key identifiers that are used to the same end across the different classes of information systems that make up these organizations' homebrew databases. For example, a last name on the top of a paper form is used for filing that form in alphabetical order then and re-accessing the data it contains later. We adopt this broader unit of analysis to enable a more holistically-informative view over work practices that are likely to inform database design—which includes not only database systems but also information systems that are used to perform similar work to that which is intended to be supported by databases. In each of these classes of information systems that make up organizations' homebrew databases, identifiers are used consistently to access records of individual clients, to aggregate records, and to deduplicate records (where possible). And in most cases, these identifiers across classes of information systems serve a significant role influencing the work of these organizations, for example:

- **The primary key, as implemented in database systems.** The HIV/AIDS focal organization receives funding from the County and State to provide a variety of health-related services. The funding is passed down from the Health Resources and Services Administration (HRSA), an agency of the U.S. Department of Health and Human Services, to the County and State which in turn contract out to local nonprofit service providers. Per provisions of the funding relationship, the organization shares client data with HRSA via a prescribed database system. A staff member describes the relative degree of anonymity afforded by the database through an explanation of how, in this instance, the *obfuscated identifier* used as a primary key is generated: "Real data is not reported by name to HRSA, it's reported using the unique identifier. And so, you know, the identifier is generated by an algorithm that looks at first name, last name, gender and date of birth" (A17).
- **Identifiers that serve as de facto, human-accessible primary keys in database systems.** Another database used within this same organizational network provides a mechanism for looking up clients' work histories via the *legal identifier* of their U.S. Social Security Number so that their "work history is then documented through your taxes and how long you worked at that employer" (A3).
- **Identifiers that serve as de facto primary keys in spreadsheets used as databases.** Three emergency housing services organizations in the homelessness case have undertaken a grassroots effort to merge their client data to try and better understand clients' use of services across organizations in the county. Once a year, each organization exports its data to Excel from the different systems each uses as their primary client database—one from Access, one from Salesforce, and one from Excel. One executive director, then, manually merges and deduplicates all records using the clients' names as *aliased or not necessarily unique, legal or pseudonymous identifiers*. A staff member at an HIV/AIDS organization manages a similar process, further specifying the workflow with *legal identifiers* to involve: "I do a sort based on names and I deduplicate" (A8).
- **Identifiers that serve as de facto primary keys in paper forms used as analog proxies for databases.** In the human services context, one of the most common paper-based "databases" are the forms used to refer clients for services at other organizations. The HIV/AIDS focal organization uses a paper referral for numerous services: "The mental health referrals, substance abuse referrals, legal referrals and referrals for medical case management, those are all the paper referrals. It's a 2-page referral so the first page is basic demographic information: client's name, date of birth, Social [Security Number], contact [telephone] number and what it is they need..." (A5). The lead information on page one, the data that serves the role of the primary key for sorting and accessing each form-as-record in this instance, are *legal identifiers* (e.g., Social Security Number), verified through government-issued ID cards.

#### 4.2 The Sociotechnical Work of the Primary Key

Never merely just a tool for carrying out the technical work of linking database tables, primary keys—and other identifiers used for primary key practices—enable numerous forms of critical social work, as well. In unpacking the social work enabled by the primary key, we begin here to employ more of the rhetorical elasticity of our broader analytic frame. Here, by 'primary key,' we are most often referring both to primary keys that are 'technically' primary keys as well as to identifiers that are used for primary key practices. In the human services context, the primary key enables social work related both to the individual who is identified and to the data that has been collected about that individual. Once a primary key has been created for a client, it is used by different stakeholders to enable different forms of work both within and across organizations.

*4.2.1 Counting Clients for Accountability: Often-Estimated Aggregations.* More than any other form of work supported by primary keys, informants talk about the work of *counting clients*. In the homelessness focal organization, staff explained that the count of "how many people have stayed each month" (H10) is frequently compiled and sent to the board because of their level of interest in the information and the metric's importance in driving their decision-making (H12). Similarly, in the HIV/AIDS focal organization, one staff member discussed the kinds of common data requests that she receives: "My supervisor [asks] '...how many positive tests did you have this month?'" (A9). The count of positive tests per month acts as a proxy for a count of clients who have been identified as HIV+ based on the assumption that only one test has been run per person. While this might be a valid assumption in many cases, informants also discussed instances in which clients obtained additional tests or nights at a homelessness shelter by providing different identifying information at each interaction.

As a staff member from a referral organization in the homelessness services network explains, a motivation for counting clients is to compare supply and demand: "Who do we serve and how do we want to serve them? I mean that's basically... every conversation gets down to it and it's like, you know, who gets priority? What programs [do] we put emphasis on?" (H6). Using primary keys to develop counts of people in need and comparing those counts against funding allocation allows decision-makers to reallocate resources. Underpinning this logistical question, however, is the more political question of which, and how many, clients are prioritized for services.

Notably, frontline staff rarely share accounts of counting clients for their own work. Counting clients is predominantly done for reporting upward in the organizational hierarchy and outward to boards, funders, and policymakers. In the context of this reporting, staff in both cases are fiercely protective of their clients' privacy. Even if staff used less anonymous identifiers as primary keys in information systems within the organization, they frequently de-identified their client data before sharing, switching to the use of *anonymous or obfuscated identifiers*: "I enter all the information into a big database that I send to the county quarterly. And I just de-identify it, so they just have a client identification number as opposed to a name for anonymity purposes" (H9). The acceptability of this (pseudo-)anonymization practice for the receiving parties suggests that the recipients of the data are, in general, more concerned with unique counts of clients than specific client identities.

Yet, contrary to normative assumptions in database design that a normalized, orderly, and efficient database contains one and only one unique (deduplicated) record in each table corresponding to its real-world counterpart, the normative assumption in the human services context is that primary keys are most likely to identify unique but not necessarily deduplicated records—that *aliased identifiers* are an expected and everyday reality in these organizations' databases: "...we found out after matching everyone first that although you might naively think there was maybe 5,000 or so people in the system in the course of a year, there was probably closer to 3,000 with dupes taken out, so that was a significant finding" (H12). That process of deduplication was reported to be a little like "detective work" (H10). In instances in which records are being tracked in spreadsheets, the counting process requires manual deduplication, often through sorting, for example:

A8: And then when it comes to reporting on overall number of households served, I don't need to count that person three times because they're on that spreadsheet three times, I need to count them once for that purpose.

I: Do you have a unique identifier that helps you do that quickly?

A8: It's name-based in that spreadsheet and so I do whenever I'm counting number of households, I do a sort based on names and I deduplicate.

Other informants report getting in trouble for having duplicate records for clients: "And then County catches those and smacks us on the hand for not figuring out that they're already in there because it causes dual records and then they need to verify which one is kept..." (A14).

Duplicated client records have to be assumed to exist both within and across organizations, however, for reasons beyond any single organization's control. Multiple funders of the HIV/AIDS services focal organization each require that records of service provision supported by their funding be entered into each of their separate databases, so the same client at a single organization might exist once, separately, in each funder's database. Staff in both cases also report being aware of instances in which clients presented at multiple human services organizations with different names, or *pseudonymous identifiers*, (e.g., "you could be Joe at the shelter and then go to [another city] and be Joseph and we would never know that you're the same guy" (H1)), making it difficult to aggregate data across organizations to derive an accurate client count for larger community-based efforts. As such, aggregated counts are more-than-likely assumed to be estimates.

*4.2.2 Tracking Clients for Effectiveness: Aspirations of Re-Identifiable, Longitudinal Client Histories.* The most common work that frontline staff wished their data and its primary keys would support, "the holy grail of information around people, is longitudinal data" (H1). Staff across cases most commonly referred to this work as "tracking" or "following" clients:

I wish we could *track* every single service that an individual will use his Ryan White [benefits] for and I thought, you know, if that was incorporated into our own database we could easily track that... but we just, when we talked about it together we just realized that would be just another database. (A7, emphasis added)

Our dream... we have long conversations philosophically about how do we, how do we *follow* people at the shelter? How do we follow a person? Say they've come in as an emergency client, they entered into the transition program, you know, they got kicked out or they came back, they went, they came back, got kicked out, been here 3 times or whatever, uh, which is not uncommon and, and yet it's hard to sort of follow that person. So we had this dream of like a data entry form where you could literally follow an individual all through their trek through the shelter. (H8, emphasis added)

In contrast to the work of counting clients, the work of following clients was envisioned as predominantly being useful for frontline staff. Despite the overtones of surveillance, frontline staff felt that longitudinal client data, envisioned as the "person's history" (H6), would enable them to better serve clients' needs—forming a holistic picture of what services clients have received and, implicitly, for what services they might still be eligible. Based on a similar sort of longitudinal client history that three organizations in the homelessness services case manually assembled over a year, they observed that:

Now we have this interesting thing.... We know [our high-use clients] by name now. We can alphabetize them, we can talk to them, you know, we can start convening committees to work around, you know, what are we going to do about it.... (H12)

These histories are anticipated to be able to "create efficiency" (H6) so that clients "don't have to keep getting cycled through" (H5). One staff member specifically affirmed that these longitudinal client histories would likely be "a little too tedious, I think, for most funders because they want the bigger picture" (H8).

And yet, not only are there challenges in aggregating data across multiple organizations' databases with incommensurate primary keys, there is a wicked irony here with respect to who can aggregate interorganizational, longitudinal data and who sees the on-the-ground value in those data. They are not the same populations. For funding agencies, longitudinal client histories are both feasible—and

in some cases (e.g., HRSA), a reality—given that funders are in a position to prescribe the use of a particular database, define the data schema, and stipulate the use of *anonymous or obfuscated, and deduplicated identifiers* to aggregate data while preserving some degree of client privacy. For example, one staff member from the HIV/AIDS focal organization (A17) did reflect that the HRSA database enforced the use of consistent, algorithmically-generated primary keys by any organization receiving funding from HRSA. This technical policy affords the creation of holistic client histories—at least for internal use within HRSA. But since not all of the focal organization's clients are supported with HRSA funding, not all clients have a HRSA algorithmically-generated primary key. As a result, for consistency within the focal organization, staff must use an alternate primary key across all their clients. This complicates the re-identification and re-association of the HRSA client's record with the actual client who frontline staff would want to serve. Through the process of aggregation, the link between the data and the client represented by the data has been compromised (see also [53]).

Nafus refers to anonymized, aggregated data that can't be re-identified as "dead data" [44]. Data organized and stored with primary keys to help ensure privacy in aggregation, then, also prevents the use of that data by the frontline staff who envision the value of leveraging these longitudinal client histories for delivering more effective services. For them, access to the longitudinal client histories that are aggregated by funders along with the ability to re-link the data to their specific client is still a pipe dream. And the prospect of creating their own databases to try and replicate these client histories is so daunting (e.g., "we just realized that would be just another database" (A7)) that it just doesn't happen.

*4.2.3 Verifying Client Eligibility for Accountability: Exchanging Identification for Access to Services.* Whereas the client histories supported by more commensurate primary keys that would enable the re-identification of records that frontline staff envisioned would *theoretically* help them better serve clients, the *actual* work that frontline staff described when talking about the identifiers that served as primary keys was the work of verifying client eligibility for services. Eligibility verification was typically used to control access to services, an enactment of the political or public policy question about which and how many clients to prioritize for services. For example, one staff member of a data aggregation organization reflected that the systems are...

...principally organized around collecting data for the feds. So, they're very regulative in nature. So, it's you know, who's the person you're serving? Are they eligible for what you're trying to give them? What did you give them? And that's about it. And so, you don't get a lot of information at that level that helps you in the actual management of the work. Or the effectiveness of the service that you're providing. (H13)

For this work, different variants of identifiers are used, depending on the way(s) in which data collection and sharing is or is not specified in agreements with funding agencies, particularly to demonstrate accountability [57]. For example, a staff member at the HIV/AIDS focal organization explained that one funder mandates the use of a database that requires a *legal identifier* for the provision of funded services: "I need their... name. I can't submit a test without it. I need your date of birth because that screening form that they complete, when that specimen, when you go to a doctor's office, when they... get the results, they've got to tie that result to somebody..." (A9). In this case, the requirement for a *legal identifier* may further be influenced by the nature of HIV as a life-threatening virus if left untreated.

The homelessness focal organization, in contrast, is required to submit client data to the federal HMIS, but the provision of services are not contingent on these data. The priority, then, shifts to getting someone off the street: "We don't require ID. Low bar for entry when it gets you in



tonight..." (H1). So while *pseudonymous identifiers* are perfectly acceptable, limited bed space at the shelter does compel the organization to limit the number of nights that any one individual stays:

In a very just like concrete way, we can only have 160 people here a night... They can only have 90 stays unless the person is in a program and then they have to wait until the next season... So making sure that we are keeping track of who's had how many stays, so then we're not letting people stay past that. (H10)

Because of this, the organization also attempts to collect *deduplicated identifiers*—"Right now we just, you give us a name and as long as [it is] the name you give us every time you come, that's cool, right" (H1)—though the nature of *pseudonymous identifiers* only affords *not necessarily unique identifiability*.

In both cases, different variants of identifiers serve as a form of currency that clients exchange for services; the variant of the identifier that is required most commonly depends on what funders stipulate, which varies both within and across cases.

**4.2.4 Referring Clients for Effectiveness: Coordinating Service Provision through Data Sharing.** Organizations in both cases reported using data to support the collaborative work of referring clients for services across organizations, for example: "We work with 14 different agencies... Our whole mission... is to make referrals, so we will like refer out all day long" (H6). Unlike other motivations for sharing data (i.e., counting or tracking clients), in which client anonymity is supported through the use of *anonymized or obfuscated identifiers*, client referrals across organizations necessitate a *legal identifier* so that the individual can be identified on handoff as the one for whom services have been requested. In both focal organizations, the referral is documented in one of the organization's own databases. The handoff to the referral organization is done via a paper form. But the one-way sharing of data doesn't fully support these organizations' work practices. In the homelessness services case, when case managers refer a client to another organization, they want to hear back from that organization: Did the client present there for services? What services did he receive? And yet case managers report that they don't generally hear back from the referral organization: "We don't get any information back...so it's very like: out. We rarely get anything back" (H6).

In these cases, referral forms use *legal identifiers* as primary keys, for example: "the first page is basic demographic information: client's name, date of birth, social, contact number and what it is they need..." (A5). Once the client presents at the referral organization, a new record is created in that organization's primary database with a new primary key: "If someone left here, for example, and went to another [organization], that person's record starts completely over at that place. It's in the system but they [the case managers] can't see anything..." (A22). For the service providers, then, the referral form severs the link between the two organizations' different primary keys referring to the same client. While it might be technically feasible to use *legal identifiers* to link the two organization's records, in practice, the records remain separate. In fact, this separation remains in place even when both organizations use a shared database; access-controls on database tables are often configured to prevent staff from accessing the client's records at another organization because of privacy concerns.

Without any link back to the referring organization's primary key identifying the client, it would be difficult for the referral organization to send back data about services rendered there. Yet, if the referral did contain a link back to the primary key of the referring organization, it would implicitly or explicitly contain proveniential metadata about the referral organization, which could provide unwarranted, undesirable, or illegally-disclosed information about the client. In the HIV/AIDS case, for example, A8 noted that within a database for clients experiencing homelessness, he could not identify his referred clients as originating from their HIV/AIDS organization as that would reveal the protected status of their marginalized client population. Alternately, one could rely

on a third-party, trusted organization (e.g., via a data warehouse or clearinghouse) to facilitate the two-way information exchange, but that would require that all organizations agree on which primary key to use or—more likely—that third party would create new primary keys. So in solving the problem of ensuring two-way information exchange for client referrals, that third party might be likely to recreate or exacerbate many of the same problems around aggregation that have already been reported here.

## 5 DISCUSSION: INFRASTRUCTURE PROBLEMS & THE COERCIVENESS OF THE PRIMARY KEY

Edwards et al. [24] argue that the attempts of researchers to create more human-centered experiences with computing systems are frequently stymied by problems with infrastructure. One class of infrastructure problems are those termed *interjected abstractions*—problems in which "low-level infrastructural concepts become part of the conceptual model of the interface" and are "exposed to users through the applications built on top of the infrastructure" [24]. That is, properties of the technical substrate bubble up to the surface where they shape the user experience. These interjected abstractions, then, force end-users to work on the infrastructure's terms. Through the interrogation of the work practices supported by variants of the primary key, we find three infrastructure problems that derive from their infrastructural properties: the immutability of identifiers, the hegemony of NOT NULL, and the demand for uniqueness within and across contexts. At the end of each discussion section, we provide provocations for design related to each infrastructural abstraction. Following these three sections, we discuss the influence of these infrastructure problems on the way that public policy is enacted.

### 5.1 The Immutability of Identifiers

In relational database design, the identifier that is used as a primary key is immutable; it simply cannot be changed. Changing the primary key would break all of the links between database tables and undermine the fundamental premise of relational databases. But some of the identifiers used as primary keys or used to generate primary keys in human services provision can and do change. For example, the database that organizations are required to use if they receive money from HRSA uses an *obfuscated identifier* as the primary key that concatenates gender and name into its identifier—both of which are mutable human attributes. Or a client who gets tested for HIV using a *pseudonymous identifier* as primary key has to get re-tested before he can receive any other services, because a client needs to have a record with a *legal identifier* verified by government identification as a primary key—and the *pseudonymous identifier* used previously as the primary key cannot be updated.

The immutability of the primary key as an infrastructural abstraction also seems to implicate how work practices around identifiers, more generally, are structured. Informants report that numerous identifiers are also treated as immutable, even though the design of the relational database would not necessarily require it. In nearly all instances, frontline staff members were not able to change key identifiers such as the "spelling of the name, date of birth, or the sex" (A14). If a frontline staff member enters that data incorrectly or if that data changes, they "have to call... and fess up" (A14). A data manager working with one of the HIV/AIDS organization's funder's databases concurs and confirms that, in practice, identifiers need to be changed "a fair amount... a couple times a week." Informants report that changing identifying information cannot typically be done by escalating the issue within the organization. Requesting changes required contacting a government funding agency (e.g., "the County" (A14)), the database administrator (e.g., HMIS (H10)), or even the Department of Motor Vehicles (A14).

**5.1.1 Provocations for Design.** While changing a database primary key is generally not recommended due to potential impacts on data integrity and database performance, it is feasible that relational databases might support ways to accomplish such a change, or at least support an alternate way of updating this information over time, such as architecting schemas that would allow for additional or alternative identifiers to be added as trust is built. That staff members have to "fess up" to incorrect identifier data also seems disproportionately onerous and punitive given the commonality of the occurrence and the inherently dynamic nature of the characteristics upon which some keys are based. Moreover, identifying instances of incorrect data and correcting that information should be affirmed and made easier to accomplish to increase the overall validity of data.

Keeping up with the changing nature of human data is a more general problem, as Raley notes, for example: "the errors inherent within a catalog mailing list, one of the more basic data sets, indicates how unstable that data can be: any given population is a massive moving target..." [51]. The infrastructural abstraction of immutability limits the ability of the database to dynamically represent a changing phenomenon. Because we see instances in which a variety of identifying information might be used in the algorithmic generation of a primary key (e.g., via an *obfuscated identifier*), finding ways to better support updating data across many different fields would likely be valuable.

## 5.2 The Hegemony of NOT NULL

The workflow imposed by relational database design is that the primary key must be established first—whether automatically generated by the system (e.g., by an algorithm) or entered by a staff member or volunteer. It is simply not possible to commit a row (e.g., add a new client) to a relational database without a primary key; fields that are assigned to be the primary key are prohibited from taking on NULL or unassigned values. This normative characteristic of relational databases also shapes data collection practices. Primary key fields or identifiers that are used to generate a primary key (e.g., in the case of *obfuscated identifiers*) are often collected before any others, whether data is collected on paper, in spreadsheets or using an enterprise database.

Frontline staff members that work in these stigmatized policy fields, and, as they report, their clients, are often wary of sharing *legal identifiers* when they believe that information will be shared with untrusted parties. As a result, frontline staff report working carefully to build rapport and develop trust to facilitate the collection of the type of highly personal information that typically serves as a primary key (e.g., legal names or governmental ID numbers) and, therefore, is problematically requested before other information:

There's a very large sense of trust that is needed between the client and myself in this position and I think that the rapport built with the client is done in a very calculated way and a very transparent way: 'Also like just so you know up front this is what I'm going to need from you...'. There are a number of very personal things that we have to focus on and there's also trust issues, too, 'cause you're asking for like social security number and all this other stuff and we're working with clientele that doesn't necessarily trust the system or authority figures or people in general. (H11)

Sometimes, as reported above, the requirement that frontline staff collect data up front that includes highly personal information serving in the role of primary key requires significant additional confidence-building work. One staff member in the homelessness services case even reports a sense that data collection has gotten even more personal and intense over time: "It's gotten quite, maybe interesting is the word. I don't know but like we want to know everybody's sexuality. We want to know everybody's this, everybody's that..." (H8). When that highly personal information

becomes a precondition for determining eligibility and/or receiving services, other challenges arise, such as the proliferation of *pseudonymous, unduplicated identifiers* when clients provide different identifiers to different, collaborating organizations which creates challenges for counting clients or merging longitudinal client histories. In other instances, frontline staff report not wanting the requirement to collect highly personal identifiers to foreclose connecting the client with services. In these instances, they accept *pseudonymous identifiers* in place of *legal identifiers*, a practice which can also create other problems down the road.

The abstraction of the primary key, specifically that it is prevented from containing a NULL value, has contributed to influencing the data collection experience, most often imposing a workflow that frontloads the need to collect the most personal information first from human services clients. The multiple workarounds created—whether the additional trust-building work before the request for *legal identifiers* or whether the collection of *pseudonymous identifiers* in their place—complicate and reinforce existing challenges in counting, tracking, determining eligibility, and making referrals.

*5.2.1 Provocations for Design.* The workflow imposed by the technical abstraction that the primary key cannot be NULL does suggest a few provocations for design. What would it mean, for example, if more flexible workflows were supported that allowed a client record to be created with a temporary or placeholder name so that whatever primary key is used by the database could be generated? Then, once a trusted relationship had been built, a legal name could be entered and the primary key could be updated across the database or appended to an existing record. This approach would ensure the orderliness of the database(s) while respecting the process of relationship-building and the desire of frontline staff to move quickly to provide services when the situation is warranted.

As discussed previously, updating primary keys is challenging from a technical perspective. However, even if it were technically possible to incorporate more workflow flexibility into the database, the infrastructural abstraction that the primary key cannot be NULL is still extremely complicated to navigate for non-technical reasons. Not all records in human services are stored digitally, and updating clients' identifiers across submitted forms or case notes is virtually impossible, even if the task of updating that information was deemed to be worth the investment of time. With many organizations required to perform duplicate or triplicate entry into a variety of internal databases and spreadsheets, in addition to external databases, updating information across each of these systems would likely be deemed to yield a questionable return on investment. While, perhaps, data warehouses or federated databases might seem to offer some promise, software vendors undoubtedly also feel a lack of return on time invested, since, with probably very few exceptions, such features are likely not being requested or used by their customers. As one informant explained,

I think the naive response to most things is: Well, we just ought to have a master system and I think that most people who have an operation in process have legacy systems and so I think there's a lot of tension between why can't we figure out ways to use our legacy systems? And, well, it's very frustrating to have to ask a question and deal with legacy system boundaries, so I think you all should just move to a new system and there's probably something that does everything, but you know there never is. So I think that is probably the place where most things stall is in that area... (H3).

### 5.3 The Demand for Uniqueness Within and Across Contexts

Another requirement in designing a relational database is that the primary key must be unique—that is, within the table of client records, each primary key must appear at most once. It is then assumed that all references to a particular primary key refer to the same entity (e.g., the client record) throughout the database. There is an important distinction between the technical requirement for uniqueness among records and the lack of a technical requirement for uniqueness among

people represented by those records. While the normative assumption in database design is that the uniqueness of the primary key should result in one-to-one mappings between unique primary keys and unique individuals in the real world, the technical requirements do not actually enforce this mapping. Instead, the properties of the infrastructure merely constrain the database in representing relationships between *some* person and the data related to that person. And in the human services context, this merely constrained criteria for uniqueness results in a proliferation of duplicated records. In order to construct a one-to-one mapping between a unique primary key and a unique individual, the individual either needs to (or needs to want to) be able to successfully connect themselves to their single identifier consistently over time by knowing their specific primary key (e.g., a U.S. Social Security Number) or by providing identifying information that allows that primary key to be looked up (e.g., one's associated demographic data like a name and/or date of birth).

While the technical requirement of the uniqueness of the primary key creates space for the duplication of records that is helpful in some ways—for the negotiation of trust or the expression of a lack thereof—it does create other challenges for human services work, particularly with respect to the duplication of identifiers and, as we turn to next, the merging of records across programs and organizations. Understanding the genesis of the many-to-one mapping between the multiple primary keys associated with an individual client is essential for understanding the additional challenges with respect to the uniqueness that become layered onto this mapping.

Given the infrastructural requirement that a primary key must be unique *within its given context*, it also bears unpacking the role of context in that uniqueness. The variants of identifiers that are allowable as primary keys, then, are selected to be those most suited to the sociotechnical demands of each context. And it is the varied needs of each context that create additional proliferation in the one-to-many mapping between individuals and their associated primary keys—this time because of the different (sometimes incommensurate) variants of identifiers that are used in each context. For example, one organization in the homelessness services case is morally opposed to verifying *legal identifiers* when they feel their job ought to be getting people into the shelter, off the streets, and out of the cold. One smaller organization in that same case explained that "we really don't want to be in possession of people's [Social Security Numbers], so we don't do that" (H12). In contrast, the focal organization in the HIV/AIDS services case is *required* to verify and collect that very same information:

We had a client who—and this happens more frequently—that they give false names when they go to get tested for HIV. And she had done an intake with a person, the name that he had given... for testing was in the false name that he gave. So she had to redo the intake under the correct name because there's no way to, because that information was null and void because of the fake name (A2).

Different contexts both within and across organizations have different demands and philosophical orientations to human services provision that create variation across the kinds of primary keys that are used.

Across databases used both within an organization (often managed by different funders, for example) and across different organizations (who may want to collaborate through the aggregation or sharing of data), the different variants of identifiers used as primary keys—the variants selected as being appropriate *within a context* but that differ *across contexts*—create even more challenges. For example, within the context of one HIV/AIDS service organization's database, identifiable primary keys are used productively; however, once the context expands out beyond the organization's walls and across organizations, the mere association between the identifiable client and the organization (e.g., via proveniential metadata) discloses a person's HIV-positive status, which is a violation of the

client's privacy. As A8 explains, "just by being affiliated with [our organization] in that database, someone could assume or make an assumption that [the client is] HIV positive."

While either of these problems might conceivably be addressed through intermediary translation tables that match and merge client records across databases—perhaps in a way that private information (such as protected health status) is limited to a small group of database administrators—the issue becomes that there are an exponential combination of contexts, each with unique work and privacy requirements, and each with a proliferation of identifiers and primary keys:

Since we use so many different databases but we also use different client IDs for clients even within medical services like, so we have our care coordination database, CaseManager, and then we have [another database], ACAP, that we use for medical services eligibility here.... And in each of those systems, half of those clients that are in care coordination are also going to be on medical services but they're going to have a CaseManager ID and they're going to have an ACAP ID. And the same across surveillance and prevention... so surveillance, you know, they use... some kind of unique, unique identification code for each person instead of a social [security number] so it's just trying to match between the programs and the different client IDs and that sort of thing. It's difficult. (A19)

As the context expands out beyond any one database or any one organization to the city, state, or national levels of scale, the lack of a consistent identifier forecloses any possibility of uniqueness. As data is aggregated at different levels, the lack of primary key consistency across contexts and over time is at odds with the infrastructural abstraction.

**5.3.1 Provocations for Design.** The entangled requirements and expectations around different manifestations of uniqueness—the infrastructural uniqueness of the primary key and the more-or-less unique mapping of those keys to an individual client—is most problematic in human services when the variety of contexts in which uniqueness must be achieved span multiple databases and multiple organizations. The overlaying of these different manifestations of uniqueness in the human services context results in a working operationalization of uniqueness as *primary key consistency* or *commensurability across contexts*, which, as we have illustrated above, is a sociotechnically difficult problem to solve.

If the problem of entangled uniqueness were solved, such as it were, it would enable forms of human services work that require deduplication but that do not demand identifiability (e.g., aggregated client counts). But other forms of human services work demand identifiability (e.g., longitudinal client histories). Here, the overlaid requirements of entangled uniqueness and identifiability begin to more starkly conflict.

The process of deduplicating and merging records that would constitute a longitudinal client history have been characterized elsewhere as a process of assembling "flecks of identity" [27] across databases. Raley refers to the resulting corpus of data that is assembled about an individual as a "data double"—an institutionally-oriented digital stand-in for an individual person that is built by merging discrete data streams together, which can then act—and be acted upon—as if it was the human being him/herself [51]. These data doubles come to represent clients in public policy conversations, organizational decision-making, and program management.

For the construction of the data double, the degree of anonymity of the primary key (i.e., anonymous, pseudonymous, obfuscated, legal) matters less for merging data into a single data double than it does for linking that data double to a unique individual in the real world. For example, records across multiple databases could be linked together to form a single data double if only *anonymous identifiers* were used, assuming they were used consistently across contexts and over time. However, without some type of translation table to connect the data double back to the



real person, this data double would be detached from the individual and largely unusable by the frontline staff who are in the best position to use these data to help clients. As identifiers increase in identifiability, it becomes easier to link the data double to a unique individual.

To accommodate this linkage, however, consistently used primary keys must also be consistently highly identifiable. This means that in order to produce deduplicated client data doubles across contexts, there would be a need to use the most invasive, least privacy-protecting identifiers to resolve ambiguities across databases. In the process of linking data to create data doubles, then, whatever privacy advantages might have been gained by using more privacy-preserving versions of primary keys would be lost. As a result, the alignment of client privacy and welfare with an appropriate degree of anonymity in primary keys should be a pivotal consideration in any specific context.

In both the HIV/AIDS and homelessness cases, different third-party organizations were talking about or more actively working towards developing a single, centralized, data warehouse and simultaneously working to establish buy-in for such an endeavor. There was both progress towards and pushback against a centralized data warehouse in both cases. The sociotechnical tensions that have been uncovered in this research do suggest, however, that initiatives to collapse the diversity of philosophies and requirements of different human services contexts into one global, unified identifier will face significant challenges. There will remain a need for different variants of primary keys, both in terms of the degrees of anonymity and uniqueness afforded, in different organizational contexts and for different forms of human services work. The variants of identifiers and the duplication of identifiers that clients help to seed across contexts that we see in this research are also consistent with how sociologists understand the multifacetedness of identity and its resistance to being collapsed into a single, unified entity [26]. Instead of advocating for centralized, unified identifiers that remain immutable, can never be null, and are globally unique, perhaps it is possible to design primary keys that embrace the complexity of identity and its changing nature within and across varied contexts.

#### 5.4 Enacting Public Policy through Infrastructural Abstractions

In the human services context, in which the pressure for organizations to become more data-driven is increasingly intense—coming from the federal government [32], from other funders [4, 9, 73], and even from cultural narratives of technological progress perpetuated by technology companies [29]—data work is not optional. But rather than being ‘data-driven,’ Bopp et al. [9] found that organizations in the social sector are, instead, being ‘driven by’ their data, and others have questioned whether this data work may even “come at great cost to themselves and ultimately to the people they serve” [58]. Along with being driven by the demands of data, then, we also find that organizations and their stakeholders are also driven by the demands of the infrastructure—often required and prescribed by funders—that are used to manage those data.

The well-established politics of classification systems [12] attunes us to the influence wielded by the categories of data that human services organizations are required to collect about their clients and the services that they provide. But here we see the influence not just of the categories encoded in a schema, but also the influence of the infrastructural abstractions underlying the material form—the relational database—in which that schema is instantiated. The materiality of the relational database [21] dictates the use of a primary key, and the infrastructural abstractions of that primary key—that primary keys are immutable, can never be null, and are globally unique—in turn, force the sociotechnical work of human services provision to be carried out in its image:

- **Infrastructural abstractions influence the nature and order of the work.** Frontline staff members are tasked with a workflow in which highly sensitive personal data are

supposed to be collected before other data—or even before services can be rendered. Frontline staff report that instead of feeling that their job is about helping clients succeed in programs, it is now just data work.

- **Infrastructural abstractions create new forms of work as stakeholders work around the constraints of the abstractions.** Frontline staff have to call to “fess up” to funders when client data changes. Testers have to add re-testing to their workload for clients who originally presented with a *pseudonymous identifier*. Frontline staff engage in additional trust-building work before requesting identifiers that could be perceived as too personal.
- **Infrastructural abstractions influence the tenor of the relationships among individuals.** Frontline staff provide accounts suggesting a disempowering relationship with those who manage and gatekeep data in databases. They also report evidence suggesting that tenuous relationships with clients are dynamically negotiated through the data-sharing process.

Edwards et al. [24] suggest that interjected abstractions from the underlying infrastructure can lead to problems in usability and usefulness at the interface layer. Here, we find that these infrastructural abstractions have a much more coercive reach, forcing people to work in the image of the infrastructure. The relational database and its infrastructural abstractions, in some ways, *becomes* the work of human services provision: “A lot of my clients get to the point now where I’m not even asking the questions. I just say, ‘hey how’s it going?’ and they’re answering everything” (A5). But even more, the relational database and its infrastructural abstractions become the *de facto* public policy. Staff members relate to clients through the data they have to collect (see also [59]), but also, as we see here, *how they have to collect it*—where the *how* is dictated in no small part by the constraints of infrastructural abstractions. When staff relate to clients through data and information infrastructures, they do so instead of relating to clients by considering their unique situation and the overall objective of the policy: to reduce the incidence of HIV/AIDS or homelessness.

Policy, then, is not merely enacted through the classification systems and systems of measurement that are directly or indirectly prescribed by legislation produced by politicians or by contractual language drafted by funders [59]. Public policy is also enacted—likely unknowingly—by software engineers who implement and maintain relational database platforms and instantiate the rules and constraints in the infrastructure that becomes coercive to the work of human services provision.

Granted, designing for human services provision will never be an entirely technical pursuit, even given contemporary demands that these organizations be increasingly data-driven [9, 29, 32]. But the diversity of values and logics that underpin computer-supported cooperative human services provision (see also [70]), in this research, suggest that infrastructural abstractions supporting more flexibility in work would be beneficial. These flexibilities would give staff the space to negotiate the layered-ness of human relationships in this context—relationships among governments, funders, and organizations; among collaborating organizations; and between clients and each of these.

## 6 CONCLUSION

Information infrastructures are, indeed, performative. The schemas and classification systems that privilege the collection and use of some data over others influence what is possible. But the identifiers and primary keys that enable the linking and accessing of those data are also performative. The technical properties that have been baked into the primary key shape work in often-coercive ways, forcing end-users to operate in the image of the infrastructure and to find ways of working around that infrastructure. In this research, then, we make the following contributions:

- Elaborate a design space for identifiers, including empirically-observed variants of identifiers along axes of anonymity and uniqueness;

- Present a descriptive account of four classes of sociotechnical work afforded by primary keys and other identifiers that do the work of primary keys;
- Identify three technical abstractions of the primary key—that it cannot be changed, cannot be empty, and must be unique—the properties of which have forced end-users to work in the image of the infrastructure;
- Suggest design provocations for better supporting identification across a variety of contexts; and
- Provide an empirical account of infrastructural abstractions coercively influencing work practice and enacting de facto public policy.

The data collected by human services organizations characterize some of society's most pressing social problems. These data provide accounts of the interdependence of various human challenges; chronicle the successes and failures of potential remedies and solutions; and support data-driven decision-making for individuals, organizations, and larger communities. If the identifiers and primary keys that hold information infrastructures together and that bring information infrastructures together were more supportive of the sociotechnical context of these critical data, we could engender more transformative outcomes for human services organizations, their clients, the communities they work in, the donors who fund them, as well as the policymakers and citizens who have a stake in these organizations delivering effective services.

## REFERENCES

- [1] Jo Bates, Yu-Wei Lin, and Paula Goodale. 2016. Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society* 3, 2 (2016). <https://doi.org/10.1177/2053951716654502>
- [2] Lehn M. Benjamin. 2008. Account space: How accountability requirements shape nonprofit practice. *Nonprofit and Voluntary Sector Quarterly* 37, 2 (2008), 201–223. <https://doi.org/10.1177/0899764007301288>
- [3] Lehn M. Benjamin. 2008. Bearing More Risk for Results: Performance Accountability and Nonprofit Relational Work. *Administration & Society* 39, 8 (2008), 959–983. <https://doi.org/10.1177/0095399707309357>
- [4] Lehn M. Benjamin. 2012. Nonprofit Organizations and Outcome Measurement: From Tracking Program Activities to Focusing on Frontline Work. *American Journal of Evaluation* 33, 3 (2012), 431–447. <https://doi.org/10.1177/1098214012440496>
- [5] Lehn M. Benjamin, Amy Volda, and Chris Bopp. 2018. Policy Fields, Data Systems, and the Performance of Nonprofit Human Service Organizations. *Human Service Organizations: Management, Leadership & Governance* (2018), 1–20. <https://doi.org/10.1080/23303131.2017.1422072>
- [6] Marc Berg. 1997. Of Forms, Containers, and the Electronic Medical Record: Some Tools for a Sociology of the Formal. *Science, Technology, & Human Values* 22, 4 (1997), 403–433. <https://doi.org/10.1177/016224399702200401>
- [7] Matthew J. Bietz and Charlotte P. Lee. 2009. Collaboration in metagenomics: Sequence databases and the organization of scientific work. In *Proceedings of the 2009 European Conference on Computer Supported Cooperative Work*. Springer, 243–262.
- [8] Pernille Bjørn and Ellen Balka. 2007. Health care categories have politics too: Unpacking the managerial agendas of electronic triage systems. In *Proceedings of the 2007 European Conference on Computer Supported Cooperative Work*, Liam J. Bannon, Ina Wagner, Carl Gutwin, Richard H. R. Harper, and Kjeld Schmidt (Eds.). Springer, London, 371–390.
- [9] Chris Bopp, Ellie Harmon, and Amy Volda. 2017. Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3608–3619. <https://doi.org/10.1145/3025453.3025694>
- [10] Nikolaj Gandrup Borchorst and Susanne Bødker. 2011. "You probably shouldn't give them too much information"—Supporting Citizen-Government Collaboration. In *Proceedings of the 2011 European Conference on Computer Supported Cooperative Work*. Springer, 173–192.
- [11] Geoffrey C. Bowker. 2000. Biodiversity Datadiversity. *Social Studies of Science* 30, 5 (2000), 643–683.
- [12] Geoffrey C. Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA, USA.
- [13] Eleanor Burt and John Taylor. 2000. Information and Communication Technologies: Reshaping Voluntary Organizations? *Nonprofit Management and Leadership* 11, 2 (2000), 131–143. <https://doi.org/10.1002/nml.11201>
- [14] Eleanor Burt and John Taylor. 2003. New Technologies, Embedded Values, and Strategic Change: Evidence From the U.K. Voluntary Sector. *Nonprofit and Voluntary Sector Quarterly* 32, 1 (2003), 115–127. <https://doi.org/10.1177/>

0899764002250009

- [15] Joanne G. Carman. 2007. Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation* 28, 1 (2007), 60–75.
- [16] Sarah Carnochan, Mark Samples, Michael Myers, and Michael J. Austin. 2014. Performance measurement challenges in nonprofit human service organizations. *Nonprofit and Voluntary Sector Quarterly* 43, 6 (2014), 1014–1032.
- [17] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications.
- [18] James Cutt and Vic Murray. 2000. *Accountability and Effectiveness Evaluation in Nonprofit Organizations*. Routledge.
- [19] Jasper De Witte, Anja Declercq, and Koen Hermans. 2015. Street-level strategies of child welfare social workers in Flanders: The use of electronic client records in practice. *British Journal of Social Work* 46, 5 (2015), 1249–1265.
- [20] Drew DeSilver. 2017. What does the federal government spend your tax dollars on? Social insurance programs, mostly. <http://pewrsr.ch/2oUsg8h>
- [21] Paul Dourish. 2017. *The stuff of bits: An essay on the materialities of information*. MIT Press.
- [22] Paul Dourish and Melissa Mazmanian. 2012. Media as material: Information representations as material foundations for organizational practice. In *Perspectives on process organization studies: How matter matters: Objects, artifacts, and materiality in organization studies.*, Paul Carlile, Davide Nicolini, Ann Langley, and Haridmos Tsoukas (Eds.). Oxford University Press.
- [23] Alnoor Ebrahim. 2003. *NGOs and organizational change: Discourse, reporting, and learning*. Cambridge University Press.
- [24] W. Keith Edwards, Mark W. Newman, and Erika Shehan Poole. 2010. The Infrastructure Problem in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 423–432. <https://doi.org/10.1145/1753326.1753390>
- [25] Sheena Erete and Jennifer O. Burrell. 2017. Empowered Participation: How Citizens Use Technology in Local Governance. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2307–2319. <https://doi.org/10.1145/3025453.3025996>
- [26] Shelly D. Farnham and Elizabeth F. Churchill. 2011. Faceted Identity, Faceted Lives: Social and Technical Issues with Being Yourself Online. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 359–368. <https://doi.org/10.1145/1958824.1958880>
- [27] Matthew Fuller. 2005. *Media Ecologies: Materialist energies in Art and Technoculture*. MIT Press.
- [28] Christine A. Halverson. 2002. Activity theory and distributed cognition: Or what does CSCW need to DO with theories? *Computer Supported Cooperative Work (CSCW)* 11, 1-2 (2002), 243–267.
- [29] Ellie Harmon, Chris Bopp, and Amy Volda. 2017. The Design Fictions of Philanthropic IT: Stuck Between an Imperfect Present and an Impossible Future. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 7015–7028. <https://doi.org/10.1145/3025453.3025650>
- [30] Yeheskel Hasenfeld (Ed.). 2010. *Human services as complex organizations* (2 ed.). Sage Publications, Thousand Oaks, CA.
- [31] Yeheskel Hasenfeld and Richard A. English. 1974. *Human service organizations: A book of readings*. Univ of Michigan Press.
- [32] Ron Haskins and Jon Baron. 2011. Building the connection between policy and evidence: The Obama evidence-based initiatives. *Using Evidence to Improve Social Policy and Practice* 25 (10 2011), 25–51.
- [33] Christopher A. Le Dantec and W. Keith Edwards. 2008. The View from the Trenches: Organization, Power, and Technology at Two Nonprofit Homeless Outreach Centers. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, New York, NY, USA, 589–598. <https://doi.org/10.1145/1460563.1460656>
- [34] Christopher A. Le Dantec and W. Keith Edwards. 2010. Across Boundaries of Influence and Accountability: The Multiple Scales of Public Sector Information Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 113–122. <https://doi.org/10.1145/1753326.1753345>
- [35] Heather MacIndoe and Emily Barman. 2013. How Organizational Stakeholders Shape Performance Measurement in Nonprofits: Exploring a Multidimensional Measure. *Nonprofit and Voluntary Sector Quarterly* 42, 4 (2013), 716–738. <https://doi.org/10.1177/0899764012444351>
- [36] Lev Manovich. 2002. *The Language of New Media*. MIT Press.
- [37] Aryn Martin and Michael Lynch. 2009. Counting things and people: The practices and politics of counting. *Social Problems* 56, 2 (2009), 243–266.
- [38] Nan L. Maxwell, Dana Rotz, and Christina Garcia. 2016. Data and decision making: Same organization, different perceptions; different organizations, different perceptions. *American Journal of Evaluation* 37, 4 (2016), 463–485.
- [39] Brenda McPhail, Terry Costantino, David Bruckmann, Ross Barclay, and Andrew Clement. 1998. CAVEAT exemplar: Participatory design in a non-profit volunteer organisation. *Computer Supported Cooperative Work (CSCW)* 7, 3-4 (1998), 223–241.

- [40] Cecelia Merkel, Umer Farooq, Lu Xiao, Craig Ganoe, Mary Beth Rosson, and John M. Carroll. 2007. Managing Technology Use and Learning in Nonprofit Community Organizations: Methodological Challenges and Opportunities. In *Proceedings of the 2007 Symposium on Computer Human Interaction for the Management of Information Technology (CHIMIT '07)*. ACM, New York, NY, USA, Article 8. <https://doi.org/10.1145/1234772.1234783>
- [41] Cecelia B. Merkel, Lu Xiao, Umer Farooq, Craig H. Ganoe, Roderick Lee, John M. Carroll, and Mary Beth Rosson. 2004. Participatory Design in Community Computing Contexts: Tales from the Field. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices - Volume 1 (PDC 04)*. ACM, New York, NY, USA, 1–10. <https://doi.org/10.1145/1011870.1011872>
- [42] Annemarie Mol. 2002. Cutting surgeons, walking patients: Some complexities involved in comparing. *Complexities: social studies of knowledge practices* (2002), 218–257.
- [43] Naja Holten Møller and Pernille Bjørn. 2011. Layers in sorting practices: Sorting out patients with potential cancer. *Computer Supported Cooperative Work (CSCW)* 20, 3 (2011), 123–153.
- [44] Dawn Nafus. 2014. Stuck data, dead data, and disloyal data: The stops and starts in making numbers into social practices. *Distinktion: Scandinavian Journal of Social Theory* 15, 2 (2014), 208–222.
- [45] Diane M. Nelson. 2015. *Who Counts?: The Mathematics of Death and Life After Genocide*. Duke University Press.
- [46] Wanda J. Orlikowski. 2010. The sociomateriality of organisational life: considering technology in management research. *Cambridge Journal of Economics* 34, 1 (1 2010), 125–141. <https://doi.org/10.1093/cje/bep058>
- [47] Michael Quinn Patton. 2006. *Qualitative Research and Evaluation Methods*. Sage.
- [48] Sarah L. Pettijohn and Elizabeth T. Boris. 2013. *Nonprofit-Government Contracts and Grants: Findings from the 2013 National Survey*. Technical Report. The Urban Institute.
- [49] Kathleen H. Pine and Max Liboiron. 2015. The Politics of Measurement and Action. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3147–3156. <https://doi.org/10.1145/2702123.2702298>
- [50] Charles C. Ragin and Lisa M. Amoroso. 2011. *Constructing Social Science*. Thousand Oaks: Sage Publications.
- [51] Rita Raley. 2013. Dataveillance and Counterveillance. In *"Raw data" is an Oxymoron*, Lisa Gitelman (Ed.). Chapter 7, 121–146.
- [52] Dave Randall, Rob Procter, Yuwei Lin, Meik Poschen, Wes Sharrock, and Robert Stevens. 2011. Distributed ontology building as practical work. *International Journal of Human-Computer Studies* 69, 4 (2011), 220–233.
- [53] David Ribes and Steven J. Jackson. 2013. Data Bite Man: The Work of Sustaining a Long-Term Study. In *"Raw data" is an Oxymoron*, Lisa Gitelman (Ed.). Chapter 8, 147–166.
- [54] Monika Schneider, Daniel Brisson, and Donald Burnes. 2016. Do we really know how many are homeless?: An analysis of the point-in-time homelessness count. *Families in Society* 97, 4 (2016), 321–329.
- [55] Dick Schoech. 2010. Interoperability and the future of human services. *Journal of Technology in Human Services* 28, 1-2 (2010), 7–22.
- [56] James C. Scott. 1998. *Seeing Like a State: How certain schemes to improve the human condition have failed*. Yale University Press.
- [57] Steven Rathgeb Smith and Michael Lipsky. 1993. *Nonprofits for hire: The welfare state in the age of contracting*. Harvard University Press, Cambridge, MA, USA.
- [58] Alana Conner Snibbe. 2006. Drowning in Data. *Stanford Social Innovation Review* (2006), 38–45.
- [59] Joe Soss, Richard Fording, and Sanford F. Schram. 2011. The Organization of Discipline: From Performance Management to Perversity and Punishment. *Journal of Public Administration Research and Theory: J-PART* 21 (2011), i203–i232. <http://www.jstor.org/stable/25836107>
- [60] Tony Sparks. 2010. Broke not broken: Rights, privacy, and homelessness in Seattle. *Urban Geography* 31, 6 (2010), 842–862.
- [61] Susan Leigh Star. 1999. The ethnography of infrastructure. *American behavioral scientist* 43, 3 (1999), 377–391.
- [62] Susan Leigh Star and Karen Ruhleder. 1996. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information spaces. *Information Systems Research* 7, 1 (1996), 111–134.
- [63] Randy Stoecker. 2007. Research practices and needs of non-profit organizations in an Urban Center. *Journal Social & Social Welfare* 34 (2007), 97.
- [64] Jennifer Stoll, W. Keith Edwards, and Elizabeth D. Mynatt. 2010. Interorganizational Coordination and Awareness in a Nonprofit Ecosystem. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, New York, NY, USA, 51–60. <https://doi.org/10.1145/1718918.1718930>
- [65] Alex Taylor. 2015. After Interaction. *interactions* 22, 5 (Aug. 2015), 48–53. <https://doi.org/10.1145/2809888>
- [66] Nitya Verma and Amy Volda. 2016. On Being Actionable: Mythologies of Business Intelligence and Disconnects in Drill Downs. In *Proceedings of the 19th International Conference on Supporting Group Work (GROUP '16)*. ACM, New York, NY, USA, 325–334. <https://doi.org/10.1145/2957276.2957283>
- [67] Helen Verran. 2001. *Science and an African Logic*. University of Chicago Press.



- [68] Janet Vertesi. 2019. From Affordances to Accomplishments PowerPoint and Excel at NASA. In *DigitalSTS: A Field Guide for Science & Technology Studies*, Janet Vertesi and David Ribes (Eds.). Princeton University Press, 369–392.
- [69] Amy Volda. 2011. Shapeshifters in the Voluntary Sector: Exploring the Human-centered-computing Challenges of Nonprofit Organizations. *interactions* 18, 6 (Nov. 2011), 27–31. <https://doi.org/10.1145/2029976.2029985>
- [70] Amy Volda, Lynn Dombrowski, Gillian R. Hayes, and Melissa Mazmanian. 2014. Shared Values/Conflicting Logics: Working Around e-Government Systems. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3583–3592. <https://doi.org/10.1145/2556288.2556971>
- [71] Amy Volda, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In *Proceedings of the 2011 SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 915–924. <https://doi.org/10.1145/1978942.1979078>
- [72] Amy Volda, Ellie Harmon, and Ban Al-Ani. 2012. Bridging Between Organizations and the Public: Volunteer Coordinators' Uneasy Relationship with Social Computing. In *Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1967–1976. <https://doi.org/10.1145/2207676.2208341>
- [73] Amy Volda, Ellie Harmon, Willa Weller, Aubrey Thornsbery, Ariana Casale, Samuel Vance, Forrest Adams, Zach Hoffman, Alex Schmidt, Kevin Grimley, Luke Cox, Aubrey Neeley, and Christopher Goodyear. 2017. Competing Currencies: Designing for Politics in Units of Measurement. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 847–860. <https://doi.org/10.1145/2998181.2998209>

Received April 2019; revised June 2019; accepted August 2019