

The Work of Reuse: Birth Certificate Data and Healthcare Accountability Measurements

Kathleen H. Pine¹, Christine Wolf¹, Melissa Mazmanian¹

¹University of California, Irvine

Abstract

Data reuse – long a key focus of information studies and CSCW research on eScience – is increasingly a major issue in organizations attempting to leverage data gathered using information systems for accountability functions such as performance measurements. Carrying out organizational analytics and performance measurements for accountability typically rests on the ability to successfully reuse existing, procurable data. We present results from an ethnographic study of the practices of recording birth certificate data and related attempts to assess and improve birth certificate data in response to a new reuse of birth certificate data for measurements introduced to hold hospitals accountable for the quality of the care they are delivering. Drawing on literature on data reuse and information infrastructure, we describe the situated work that must take place in order for birth certificate data to be reused for accountability purposes, and the organizational capacity building that must take place to facilitate the reuse of birth certificate data for measurement oriented to organizational improvement and accountability.

Keywords: Information infrastructure; accountability; performance measurement; data reuse; healthcare

doi: 10.9776/16320

Copyright: Copyright is held by the authors.

Acknowledgements: The National Science Foundation Grant 1319897, the Intel Science and Technology Center for Social Computing, and the Center for Organizational Research supported this work. The researchers also thank Mary Pauline Lowry for her valuable contributions to this research.

Contact: Kathleen H. Pine: khpine@uci.edu

1 Introduction

Data reuse (taking data collected for one purpose by one entity and mobilizing it for a different purpose and/or use by a different entity) is an important area of research in information studies, CSCW, HCI, and related fields. While much research on data reuse focuses on eScience, data reuse is increasingly a topic of concern in the domain of organizational management, given the intermingling of management science with data science – the interdisciplinary field focused on extracting insights from data in various forms – that places data science tools, coupled with mandates for transparency, central to management structures in both public and private organizations. Through detailing and analyzing organizational performance everywhere from schools to hospitals to construction sites, ‘big data’ promises a means to understand and improve large-scale projects. Scholars assert that the emerging “...capacity to collect and analyze data with an unprecedented breadth and depth and scale” (Lazer et al., 2009, p. 722) will revolutionize research practices in industry, academia, and the public sector (Anderson, 2008).

While a rich body of work on data reuse exists, this work tends to focus on eScience or the reuse of data in large-scale scientific research. There is a need for research on data reuse in domains outside of eScience as tools and capacities for generation, storage, and analytics are adopted in diverse public and private domains. Digitization and the wide availability of data tools, coupled with institutional pressure for audit, have resulted in large demands for performance measurement (the process of quantifying the efficiency and effectiveness of past actions, according to Neely et. Al., 2002) and policy innovations that explicitly link performance measurements to systems of reward and sanction (Dunleavy et. al., 2006; Power, 1997). In the United States, for example, schools are regulated using an increasingly sophisticated array of quantitative metrics—based on computerized testing—that evaluate student learning and by proxy teacher and school performance (Anagnostopoulos, Rutledge, & Jacobsen, 2013). Healthcare is also grappling with new demands for data-driven accountability. For example, the U.S. Department of Health and Human Services announced just this year that by 2018 90% of regular payments from Medicare will become contingent upon measurements of the quality of care delivery. The ability to make such a sweeping policy change -- a shift from paying for ‘quantity to quality’ of care -- is directly linked to the ability to quantitatively measure “quality” using large-scale data. U.S. policy, particularly the American Recovery and Reinvestment Act, accelerated adoption of Health Information Technology such as Electronic health Record systems in the hopes that this digital infrastructure would facilitate data-driven improvement and regulation. Yet, on the ground, healthcare measure developers have been unable to capture this deep repository of data. Instead, they turn to available data sources that

are already somewhat standardized and easier to extract across a diverse population of organizations. Healthcare organizations are the midst of grappling with difficult questions of how to repurpose data collected for one use (i.e. administrative data) and repurpose it for consequential quantitative metrics.

In the information sciences, there is a need for studies of cyberinfrastructure and e-Research for public administration, management, and governance. In this vein, we empirically examine a key aspect of data reuse: the complex organizational capacity building that must take place for data reuse to occur in the face of a new demand placed on an existing data source. We assert that human dimensions such as retraining occupational groups to hold novel skills required for data quality and inter-occupational collaboration are necessary but often overlooked resources in assessments of the productive possibilities of reusing data. In particular, we use ethnographic methods to examine the data work performed by a particular occupational group – birth certificate clerks – in relation to the larger lifecycle of data for performance measurement of obstetrical care.

2 Relevant Work

Recent years have seen an intensification of interest in data and data infrastructures both within our scholarly community and beyond. An important strain of work has focused on how the organization of information shapes human action, power, and culture, though this influence is often invisible (Agre, 1994; Barocas, Hood, & Ziewitz, 2013; Bowker & Star, 1999; Gitelman, 2013; Suchman, 1993). This rich body of work, which often draws deeply from Science & Technology Studies (STS) dating back to Suchman's quintessential work (1993), critically interrogates the assumptions as well as the social and cultural foundations of data, often described and treated as a free-floating, pre-existing entity that can be harvested or mined, just as one mines a natural resource. Drawing on Foucault, Ribes and Jackson have called this belief the "commodity fictions" of data (Ribes & Jackson, 2013, p. 147). This belief assumes data are given-in-advance of discovery and are "natural" objects that can circulate and aggregate regardless of their origins. Yet, both theoretical and empirical work shows that data and its attendant processes of measurement, curation, maintenance, and use implicate and even reproduce power dynamics, knowledge systems, and culturally-based assumptions (Bowker & Star, 1999; Gitelman, 2013). Thus, data is never actually "raw" but is instead always socially and culturally situated.

Knowledge sharing and organizational memory

One stream of work on data reuse focuses on reuse as part of knowledge sharing in organizations, sometimes referred to as organizational memory. Much of this literature examines information sharing and reuse to inform design and implementation of applications to support it (see for example Harper & Sellen, 1995; Orlikowski, 1992). Such studies demonstrate that simply creating repositories of information accomplishes little. Successfully deploying organizational memory via mobilizing previously stored information rests on successfully moving between contextual data collection, de-contextualization, and re-contextualization as well as ongoing activity at both the personal and individual levels (Ackerman & Halverson, 1998).

Data reuse in eScience

Another rich stream of literature on data sharing and reuse comes from studies of eScience, where the development of cyberinfrastructure to facilitate sharing is predicated on the idea that reusing scientific data is possible and necessary to advance scientific discovery on a large scale (Faniel & Jacobsen, 2010). As Borgman states: "If the rewards of the data deluge are to be reaped, then researchers who produce those data must share them, and do so in such a way that the data are interpretable and reusable by others (2012)." This body of research focuses on the release of scientific data for use by others. To this end, studies have examined social, cultural, and technical antecedents and practices of data sharing to suggest ways in which we might better promote and support sharing and reuse (Birnholtz & Bietz, 2003). Data reuse involves collaborative, intensive, and iterative labor (i.e. Rolland & Lee, 2013) including a number of ongoing stewardship practices (Karasti, Baker, & Halkola, 2006), conditions of trust (see for example Bietz & Lee, 2009; Faniel & Jacobsen, 2010), and knowledge of local contexts in which data was created (Zimmerman, 2007).

Data reuse in e-research beyond eScience

Despite the rich body of work on social, cultural, and technical aspects of data reuse, there is a need for research on data reuse for e-research beyond eScience. E-research is increasingly employed in diverse domains such as management, education, healthcare, traffic, and governance. Organizations are attempting to parlay the newly affordable array of database technology, enterprise systems, and data

analytics software packages into e-research that will improve organizational strategy and performance (McAfee, & Brynjolfsson, 2012). Studies of data reuse in organizations provide a useful starting point for the study of data reuse in these contexts. For example, Ackerman and Halverson's (1998) study of telephone hotline operators shows that using information to create organizational 'memory' requires that individuals attend to future trajectories of information as well as downstream users of information attending to contextual aspects of the information that has come into their hands, and constrains its potential uses and interpretations. Yet, these studies do not examine cases in which organizational reuse of a particular dataset changes dramatically due to a new demand placed on the data, as is the case in the e-research now occurring in healthcare organizations that are deploying tools such as data warehousing and analytics to ask novel questions of existing data resources.

In multiple sectors, increasing public demand for transparency is driving implementation of e-research (Pentland, 2000; Power, 1997). 'Data-driven' accountability, characterized by increasingly fine-grained measurements of organizational and individual performance, combines public demand for transparency with emerging data tools and practices (Ackerman & Halverson, 1998; Pine & Mazmanian, 2015). While performance measurement bears much resemblance to the scientific activities that are the focus of eScience literature on reuse, the applied science of performance measurement has some important differences. Performance measurement deals with multiple levels of abstraction (Redden & Low, 2012), often does not adhere to statistical standards found in other disciplines (Redden & Low, 2012), and is typically carried out by specialists or managers whose primary area of expertise lies elsewhere. Given the differences in the scientific activity itself, and the institutional stakes of performance measurement, there is a need for studies of data reuse for e-research related to data-driven accountability – our work contributes empirical research on the work of data reuse in the domain of healthcare.

2.1 Data Reuse in Healthcare

A rich literature addresses documentation in medical practice (see for example classic work by Berg & Bowker, 1997; Heath & Luff, 1996). Much research has focused around documentation in healthcare organizations to inform design of information systems to support medical practice. For example, Zhou and colleagues (2010) describe how EHR systems are designed based on a conceptual model that is inherently process-oriented. As a result, valuable information about the patient's psychosocial condition and history are not recorded, which limits the potential value of reusing medical record data. Yet despite the wealth of health informatics research, there is currently a dearth of research examining data reuse in healthcare.

Data-driven accountability in U.S. healthcare

U.S. hospitals are continuously under pressure from multiple external stakeholders, including the public, the Centers for Medicaid and Medicare Services, The Joint Commission ("TJC," the largest accreditation body for healthcare organizations in the U.S.), consumer advocacy groups, and large purchasers of healthcare, to improve the safety, quality, efficiency, and effectiveness of healthcare delivery. Hospitals and clinicians find themselves subject to multiple cumbersome reporting requirements, in addition to internal data-driven quality improvement initiatives typically seen at high performing hospitals. Data-driven performance metrics are being adopted as a regulatory mechanism for aligning hospital and provider performance with up-to-date standards for evidence based "best practice" to decrease variation in healthcare service delivery (Winthereik, van der Ploeg, & Berg, 2011). In the U.S., pay-for-performance initiatives and consumer-facing ranking systems are gaining traction. Clinicians and hospitals cannot escape demands for reporting data on the quality of their services; those that fail to meet reporting requirements or to improve performance in alignment with quality standards are expected to perish.

It is against this backdrop that we examine the occupational shifts that take place surrounding an emergent demand for the reuse of birth certificate data collected by clerks working in hospital medical records departments. Our study examines documentation and reuse of healthcare data focusing around the work of a particular group of non-clinical hospital personnel: birth certificate clerks who record birth certificate data. This case is unique in comparison to existing work (i.e. work by Zhou, Ackerman, & Zheng, 2010) in that data is not being reused for future clinical work, but for the calculation of performance assessments, a portion of which are reported to external regulatory entities. Our work thus extends prior literature on data reuse in eScience, e-research beyond eScience, and documentation in healthcare through examining the work of reusing birth certificate data for performance measurement, and the organizational capacities that must be developed for this reuse to occur.

3 Methods

This study utilizes empirical data gathered using ethnographic observation and ethnographic interviews. The data includes ethnographic fieldwork in three hospital field sites. These hospitals are part of a larger hospital system in located in California, comprised of six hospitals and a number of ambulatory clinics. We conducted observations of birth certificate clerks working in the medical records departments of the three hospitals that offer obstetrical services. The three hospitals vary widely in the volume of obstetrical patients they serve. The largest hospital serves upwards of 5,000 obstetrical patients per year including very high-risk cases. The second largest hospital serves around 2,000 obstetrical patients per year, and the smallest hospital serves less than 600 patients per year. The high volume hospital employs two full time birth certificate clerks and multiple temporary clerks on an as-needed basis, the medium volume hospital employs one full time birth certificate clerk and one part time clerk for evenings and weekends, and the small hospital uses a clerk who works part time on birth certificates and part time elsewhere in the medical records department.

The researchers conducted four observations in the medium volume hospital, two in the high volume hospital, and two in the low volume hospital. In each hospital (except the low volume hospital) we observed multiple clerks. We conducted observations in the medium-volume hospital first, and used observations in the small- and large- volume hospitals to corroborate data collected in this first hospital—this is reflected in the number of observations conducted at each site. Observations totaled 40 hours and centered on birth certificate data collection or activities related to using or improving birth certificate data. We also observed three interactive online trainings for clerks and their supervisors (12 hours total). This data is supplemented by observational data (160 hours to date) collected in a statewide data center that provides a platform for reporting and feedback on obstetrical performance measures from hospitals in the western coastal states. The data center is housed in a state-funded maternal quality improvement organization.

Finally, the researchers conducted interviews with stakeholders from both the hospital system and the data center who work with birth certificate data. These interviews include the supervisors of birth certificate clerks at all three hospitals, the head of medical records for two of the hospitals, the Chief Medical Informatics and Chief Transformation officers for the healthcare system, the educator for obstetrics in the high volume hospital, the supervisor for a statewide birth certificate clerk training program, and three program officers doing work related to birth certificate data quality in the quality improvement organization. Interviews averaged 50 minutes in length, and varied between 30 minutes and 80 minutes.

The researchers used an interpretive approach to qualitative data analysis in which inductive theorizing arose directly from data (Lofland & Lofland, 2006). As the researchers collected data, the researchers shared data transcriptions and developed emerging themes. Researchers created a table of emergent themes in a word document, where we placed pertinent data from observations and interviews. The primary mode of analysis was writing memos on emergent themes related to data reuse, and reworking the researchers' understandings in light of literature and new data on an ongoing basis.

4 Case

The birth certificate has long served dual purposes: as a certificate of live birth which also triggers a number of bureaucratic processes (such as registration with the social security administration) and as a convenient opportunity for monitoring public health through data collection. The birth certificate data entry process generates a long form. The top portion is the "birth certificate" as we commonly think of it. The bottom portion detaches and is maintained in both hard and digital copy at the state's Office of Vital Records. Birth certificate clerks are responsible for both portions. They gather this data by inspecting the clinical record and conducting direct data collection with patients prior to discharge. Patients fill out a worksheet form that is distributed and collected by the clerks. Another portion of data comes directly from clinical records of clinicians (i.e. clinical notes and the delivery summary written by the physician or midwife and the labor and delivery summary completed by the patient's nurse). Clerks enter data into AVSS, the software program that transmits birth certificate data to the Office of Vital Records. Because clerks retrieve a portion of data from clinical records, some birth certificate data is already a reuse of clinical data.

Once the birth certificate data has been entered into AVSS, it is archived by state and federal vital records departments. Historically, vital records data has been available to certain parties for research. In 2012, TJC rolled out a set of obstetrical performance measurements and soon required that all hospitals serving obstetrical patients collect and report data for these measures or risk losing accreditation. While in theory the widespread implementation of EHR would promote use of clinical records data for performance measurement, in practice this data was deemed too difficult to extract by designers of the

quality measures. Thus, performance measurement algorithms were designed using data elements from a variety of sources selected to strike a balance between detail of information, ease of extraction, and data quality (i.e. standardization of data elements). Birth certificate data was identified as a source of key data elements not easily extracted from any other source.

4.1 Upstream and downstream data

Multiple orders of data creation and reuse must occur to carry out a quality measurement. Second-order data work occurs after clinical practice is complete, as coders, billers, clerks, and researchers peruse medical records and capture data for organizational and administrative purposes. Performance measurement represents a third order data use since data elements for algorithms are drawn from second-order data (i.e. administrative data). Some birth certificate data constitutes a second order reuse of data that originates from clinical data. Other birth certificate data is first order data, collected directly by clerks from patients.

When TJC rolled out the new required measurements, it quickly became apparent that birth certificate data that had previously passed muster for vital records was in no way high quality enough to stand up to the rigors of standardized quality measurements. Thus, in our three hospital field sites, birth certificate clerks were grappling with a situation where they had heretofore enjoyed a wide margin of error in accuracy and completeness of birth certificate data but were suddenly held to a higher standard. We chose to focus the present analysis on the work of birth certificate clerks because they present a case where an emergent and previously unimagined data reuse created: 1) the need for shifts in the practice of recording birth certificate data and 2) new occupational and organizational capacities enabling higher-quality data production.

5 Findings

5.1 Selection of Data Elements

Before examining the impact of the birth certificate data reuse on clerks, we wish to briefly discuss how this data was selected for reuse in the first place. A number of organizations provide guides to evaluating the quality of healthcare performance measures (see <http://www.qualitymeasures.ahrq.gov/>). A national non-governmental body, the National Quality Forum (NQF), assesses healthcare measurements in the U.S. using set criteria. Among these criteria are whether the measurement will produce reliable and valid results about care quality across a nationwide sample ('scientifically acceptable'), and whether data required to calculate a measure is 'feasible to collect.' To be economical, performance measurements are designed to use available data sources, thus measure designers hoping for NQF approval and wide adoption design measurements that use the most reliable data elements that are also extractable.

Measurement guides and official statements about criteria for measure endorsement portray the process of measurement design and criteria for its evaluation as scientific, rational, and relatively clean. However, in practice selection of data elements for quality measurements is strongly influenced by sociopolitical concerns. For example, a program officer at the quality improvement organization revealed in a conversation during observation that:

'...selection of gestational age from birth certificates was strategic. They wanted to get at early elective deliveries, because that is about improving care for moms, and the majority of maternity care improvement has focused on outcomes for babies.'

Gestational age is not recorded in ICD-9 codes, so the birth certificate is the only passable source for this data element. Designers of the early elective delivery measure lobbied the national bodies that vet quality measures to accept birth certificate data as a source of data elements, even though it was known that birth certificate data was relatively low quality data, because it was important to a maternity care advocates that they target early elective delivery as an effort focused on improving care for mothers rather than focusing exclusively on health and well being of babies.

5.2 Data Quality: A Moving Target

Once the new quality measurements were rolled out, large disparities in quality and recording of birth certificate data related to the birth certificate data elements became apparent. For example, the birth certificate lists the gestational age of the fetus at birth. Getting accurate gestational age data is challenging, yet the consequences of not recording this data accurately are large. Take for example this demonstration from a webinar conducted by the statewide data center about improving birth certificate quality:

“So it looks like the birth clerk didn’t put in the right gestational age...this hospital dropped from a 17% early elective delivery rate to a 2% elective delivery rate because of data accuracy issues.”

The trainer is pointing out that once inaccurate charts were removed from the sample, a hospital’s rate of early elective delivery dropped by 15%. The early elective delivery algorithm measures the number of elective (voluntary, not medically necessary) deliveries that took place between 37 and 39 weeks of gestation. The window of time is very narrow, and being off by even a day can put a delivery in the window where it will be counted as an early elective delivery.

The training program for birth certificate clerks to produce accurate gestational age teaches clerks to select the most accurate source of a woman’s last menstrual period and gestational age of the infant. If there is doubt about the gestational age (for example, if it is recorded differently in different places in the medical record), clerks are taught to use an estimated due date wheel, a clinical tool for dating pregnancies based on last menstrual period that is distributed to clerks as part of the training. Although all of the clerks in each of the three hospitals in our study had undergone training and were part of the same hospital system, wide variations existed between the three hospitals regarding birth certificate data recording practices. This was evident in the practices of recording multiple data elements. For example, in the low-volume hospital, the clerk did not use the estimated due date wheel to resolve ambiguous data related to gestational age, while clerks in the medium and high-volume hospitals did this regularly. The low-volume hospital generally has a reputation for producing poorer quality data, yet we have not been able to determine if this is due to organizational processes, organizational culture, or other factors. However, this hospital is smaller on the whole than the other two, and does not have the resources for dedicated quality assurance personnel or a dedicated birth certificate clerk that larger hospitals often have.

Our research revealed that not only are data reuses emergent, but pressure on hospitals and clerks to produce high quality reuse data is intermittent and hard to predict. Data quality is a moving target. Since cultivating a data element for a new reuse is resource intensive, hospitals may be slow to put resources toward this purpose until they are required to do so. Our hospital field sites feel that, as the chief of data analytics for the hospital system put it: “We are constantly playing a game of catch up.” In the case of birth certificate data, hospitals and clerks have been subject to a string of demands for higher quality data for different purposes. For example, before early elective delivery was on the radar, the state’s public health agency ran a campaign to get clerks to more accurately record diabetes on birth certificates because they wanted to preserve funding for a diabetes intervention targeting pregnant women, thus needed accurate counts of diabetes rates in the state. As other performance measurements, research initiatives, and demands for funding justification come to the fore, additional portions of the birth certificate data will likely come under scrutiny, resulting in new pressures for data reuse and development of attendant processes. Assessments of “quality” are reactive, and dependent on the demands that have triggered the assessment of a data source’s quality in the first place.

5.3 Clerks as High Stakes Data Workers

Clerical workers who work with key data sources take on new responsibility, with attendant opportunities for increased status and blame. While our empirical data focus on clerks, it is worth noting that we selected this group because they are the focus of birth certificate data quality improvement initiatives. Clinicians are secondary targets in these initiatives if they are targeted at all. This is surprising given that clerks parse much of the data recorded on birth certificates from medical records. Hospital culture is strictly hierarchical and clinicians already face mounting demands for paperwork. However, clerks find themselves in a difficult position where they were often blamed for poor quality data, whether or not the problem actually originated with their work.

The program director and trainer for a birth certificate clerk training described a situation that exemplified the bias to assume that inaccurate data collection is the fault of clerks:

“...what I found as I was going around so many administrators are sure it’s a birth recorders fault, right? ...The last one I did...the clinical nurse specialist, when I showed her that their fetal presentation didn’t match, that their delivery route didn’t match. She went through their birth records. And birth recorders say, ‘No, it’s not always there’...when the clinical nurse specialist pulled up the first chart, she actually pulled up the patient chart looking for fetal presentation, there was none in the doctors note. She goes, “It can’t be missing that often...And she just sat there and goes, “Oh my God, this is horrible!”

Not only are clerks targeted as the key to improving data quality, clerks themselves have no means to request clarification from clinicians for ambiguous or missing data. Unlike medical coders, who have an official mechanism for querying providers for clarification, there is no standard procedure for clerks to do so. While medical coding work results in billing and payment, birth certificate work does not.

Thus, clerks are lower paid, have lower educational requirements, and lower occupational status than medical coders. Their practice is also less standardized across hospitals and they do not have formalized mechanisms for coordinating with clinicians.

The director of the training program said in an interview that she believes this situation is beginning, slowly, to change. She described sitting in meetings where the head clerk sat at a table with the head of the department of obstetrics. She highlighted the fact that many clerks feel increased pressure, but don't feel they have the resources necessary to produce higher quality data; interviews with clerks supported this claim. Given the increasing importance of birth certificate data, there is a need to create new organizational arrangements in which clerks can coordinate directly with clinicians. However, this will require overcoming the hierarchical culture of hospitals and giving birth certificate clerks and their supervisors not just the means to communicate their needs but the social capital to assure they are taken seriously.

5.4 Clerks exercising clinical judgment

Birth certificate clerks are not medical professionals. The only requirement for clerks is that they hold a high school diploma. Yet, they are required to accurately collect complex clinical information from the medical record and spot inaccuracies. Thus, clerks exercise discretion that requires somewhat sophisticated medical knowledge. For example, clerks must select from among a variety of potential sources of data to determine gestational age. One clerk narrated the process in the following way:

“So you are looking...to see what it is on the worksheet. And then you look in Epic [electronic health record software]...If it is in the worksheet and it is in Epic, you go with Epic. But again, if it's like completely off, you have to go by the pregnancy wheel because...I know it's a little weird. Pregnancy wheel trumps everything now...It's one thing I hate about it.

Clerks are exercising judgment to determine the best source of data and, when they detect ambiguity, they must “wheel it out” themselves. The clerk quoted here indicates that she “hates” this situation. This is because patients read over the birth certificate data in its entirety after it has been entered, and patients sometimes notice when she has entered a last menstrual period that is different than what the patient actually reported. This puts clerks in a difficult position—one clerk reported that she feels uncomfortable pushing back and tends to go with the patient's self reported data:

“So I totally go with what the mom says at that point, you know? I mean, unless it changes that she was 48 weeks, you know? Obviously that wasn't true, you know?”

This finding unearths the difficulty that clerks face in that they exercise judgments that are, for all intents and purposes, clinical judgments, and are in fact urged through their new training to do so; yet, they do not have the status as an occupational group to lend these judgments any authority.

Initial reports from the birth certificate clerk training program indicate that the quality of birth certificate data is improving based on the training. Embedded in the training is a realization that, in the current system, the fundamental work of birth certificate clerks is changing. Birth certificate clerks must learn to become skeptical and informed users of records. Organizational advocates are needed to help them attain resources. In the large-volume hospital, the perinatal educator has become involved in the birth certificate data collection process because she has directly observed differences in results of performance measurements based on birth certificate data quality. The educator, along with other key stakeholders, has lobbied for increased attention to birth certificate data, and the hospital is currently developing a LEAN project centered directly on developing new organizational processes that will support high quality birth certificate data. It is worth noting that clerks are not receiving an increase in pay and benefits despite the fact that their scope of practice is expanding. It remains to be seen if and how the social status of clerks will change as the hospitals continue to reorganize around the need for high-quality birth certificate data that can be reused for performance measurement and other e-research.

6 Discussion and Conclusion

Exploring the problems created by attempting to repurpose an existing source of data by plugging it into a new set of algorithms reveals, once again, that data does not simply sit in ready repository. Given past CSCW research on this topic, it is not surprising that birth certificate data must be carefully re-worked to fit the demands of the performance measurement.

Our research points to the evolving nature of data “quality,” which must always be determined in relation to data usages. Classic organizational theory describes the natural desire to satisfice in each level of organizational work (March & Simon, 1958). As Bowker & Star (1999) point out, people often do the doable job, not the ideal job. This is not inherently negative—good use of organizational resources means that there is no need to do a better job than is actually necessary to meet current imperatives.

But, the calculus of “good enough” changed drastically with the emergence of a new reuse of birth certificate data. It is easy to cast blame when attempts to fit a data source to a new data reuse make data that was formerly “good enough” suddenly appear to be of poor quality. If the case of birth certificates is any indication, sourcing data elements for performance measurement or other e-research will almost inevitably uncover ‘flaws’ that previously carried little consequence and were overlooked.

How should organizations deal with this situation? Ackerman & Halverson (1998) describe how effective reuse rests on data creation, de-contextualization, and re-contextualization. Successfully navigating reuse, they argue, requires thinking ahead to future trajectories of data as well as thinking back to relevant contextual information that may constrain present uses. But, we present a case of emergent reuse, impossible to fully predict in advance. Our research suggests that attending to trajectories is indeed crucial. But, when an existing data source is tapped for a new purpose, designers, implementers, and end-users should work to design new data trajectories as well; this may require new forms of coordination between upstream and downstream data users who were not formerly connected. For example, building a querying mechanism between birth certificate clerks and clinicians would give clerks a direct line to clinicians. Creating a new process to loop clinicians directly into the birth certificate data collection process, perhaps by having clinicians fill out the birth certificate worksheet as part of their delivery paperwork, may be another option. Of course, clinicians are already quite busy, so these options are not without consequence.

As reuses emerge and new data trajectories are developed, there is a need to engage in consciousness-raising so that upstream and downstream workers understand the consequences of their own data work on other groups. For example, birth certificate education efforts have targeted birth certificate clerks but fail to target clinicians. As the performance measurements calculated using birth certificate data elements become increasingly consequential (for example, as pay for performance initiatives are rolled out over the next few years) it is likely that clinicians will become motivated to understand how measurements are calculated. Ideally, this will happen in advance of weighty sanctions, but including clinicians in data quality efforts would require both bucking hierarchical medical culture and finding ways to improve the quality of clinical data recorded by clinicians without impinging on the quality and efficiency of their care. Balancing first-order medical documentation with later-order reuses of this data without impinging on clinicians working at the sharp end of care is a topic in need of more research.

We do not wish to argue against reusing data in healthcare or other domains. Rather, our research points out that reuse requires the creation of new practices and structures, and consideration of the creation and shaping of data upstream. It is inevitable that there are frictions and failures when fitting the birth certificate data to performance measurement algorithms. Yet, the selection of birth certificate data as a data element was intentional and thoughtful; experts who were party to the construction of obstetrical accountability measurements explicitly selected birth certificate data because it would allow them to measure elective deliveries (among other potential future measurements). Yet, reusing data carries risks that cannot be taken lightly. In turning to broader trends toward e-research happening in organizations, we would like to point out that our case is one of high-value data where stakeholders were motivated to inspect the accuracy of the new measurements. This led to detection of inaccuracies and multi-organizational efforts to address the emergent problems. It is likely that in other cases data will be (or is already being) re-purposed with no careful inspection of the integrity of the data once it is put to a new use.

Finally, our research suggests that it is important to carefully examine the impacts on particular occupational groups central to data work and data reuse. The burden for improving birth certificate quality rested largely on low-wage clerks, as did the blame for existence of data flaws in the first place. Designers, implementers, and end users of data need to consider the labor that goes into improving data for emergent reuses, and how this labor may fall differentially on groups who enjoy differing degrees of social capital. Designing new organizational arrangements to produce higher quality data should also take into account the need to give critical groups the necessary resources and social capital to do the required work.

7 References

- Ackerman, M. S., & Halverson, C. (1998). Considering an organization's memory. In Proceedings of the ACM Conference on Computer Supported Cooperative Work. 39-48.
- Anagnostopoulos, D., Rutledge, S. A., Jacobsen, R. (2013). The Infrastructure of accountability: data use and the transformation of American education. Harvard University Press.

- Agre, P.E. (1994). Surveillance and capture: two models of privacy. *The Information Society*, 10: 101–127.
- Anderson, C. (2008). The end of theory, will the data deluge makes the scientific method obsolete? Edge, [Online] Available at: http://www.edge.org/3rd_culture/anderson08/anderson08_index.html (25 July 2011).
- Barocas, S., Hood, S., & Ziewitz, M. (2013). Governing algorithms: a provocation piece. *SSRN Electronic Journal*.
- Berg, M., & Bowker, G. (1997). The multiple bodies of the medical record. *The Sociological Quarterly*, 38, 3: 513-537.
- Bietz, M.J., & Lee, C.P. (2009). Collaboration in metagenomics: sequence databases and the organization of scientific work. In *Proceedings of European Computer-Supported Cooperative Work (ECSCW '09)*: 243-262.
- Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: supporting sharing in science and engineering. In *Proceedings of the Conference on Supporting Groupwork (GROUP'03)*: 339-348.
- Borgman, C.L. (2012). The conundrum of sharing research data. *JASIST*, 63, 6: 1059–1078.
- Bowker, G.C. (2008). *Memory Practices in the Sciences*. Cambridge: The MIT Press.
- Bowker, G., & Star, S.L. (1999). *Sorting Things Out: Classification and its Consequences*. MIT Press.
- Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. 2006. New public management is dead—long live digital-era governance. *Journal of Public Administration Research and Theory*, 16, 3: 467-494.
- Faniel, I. M., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *JCSCW*, 19,3-4: 355-375.
- Harper, R., & Sellen, A. (1995). Collaborative tools and the practicalities of professional work at the international monetary fund. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*: 122-129.
- Heath, C., & Luff, P. (1996). Documents and professional practice: “bad” organisational reasons for “good” clinical records. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '96)*: 354-363.
- Gitelman, L. 2013. *Raw Data Is an Oxymoron*. MIT Press.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). *Computational Social Science*, 323: 721-3.
- Lofland, J., & Lofland, L. H. (2006). *Analyzing Social Settings*. Wadsworth Publishing Company.
- Karasti, H., Baker, K.S., & Halkola, E. (2006). Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work*, 15, 4: 321-358.
- March, J. G., & Simon, H. A. (1958). *Organizations*. Oxford University Press.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*. October, 2012.
- Neely, A. D., Adams, C., & Kennerley, M. (2002). *The performance prism: The scorecard for measuring and managing business success*. London: Prentice Hall Financial Times.
- Orlikowski, W. J. (1992). Learning from notes: organizational issues in groupware implementation. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW '92)*: 362-369.
- Pentland, B. T. (2000). Will auditors take over the world? Program, technique and the verification of everything. *Accounting, Organizations and Society*, 25: 307–312.
- Pine, K., & Mazmanian, M. (2015). Emerging Insights on Building Infrastructure for Data-Driven Transparency and Accountability of Organizations. *Proc. iConference 2015*.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.
- Redden, G., & Low, R. (2012). My School, education, and cultures of rating and ranking. *Review of Education, Pedagogy, and Cultural Studies*, 34, 1-2: 35-48.
- Ribes, D., & Jackson, S.J. (2013). Data bite man: the work of sustaining a long-term study.” Lisa Gitelman (ed.). In *“Raw Data” Is an Oxymoron*, Cambridge, MIT Press: 147–66.
- Rolland, B., & Lee, C. P. (2013). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. *Proceedings Conference on Computer-Supported Cooperative Work (CSCW '13)*, 435-444.
- Suchman, Lucy. (1993). “Do Categories Have Politics?” *Computer Supported Cooperative Work*, 2, 3: 177–190.

- Winthereik, B.R., van der Ploeg, I., & Berg, M. (2011). The electronic patient record as a meaningful audit tool: accountability and autonomy in general practitioner work. *Science, Technology, & Human Values*, 32, 1: 6-25.
- Zhou, X., Ackerman, M., & Zheng, K. (2010). Doctors and psychosocial information: records and reuse in inpatient care. *Proceedings of the Conference on Human Factors in Computing Systems (CHI '10)*: 1767-1776.
- Zimmerman, A. S. (2007). New knowledge from old data the role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values*, 33, 5: 631-652.