Designing for Researcher Access in the U.S. Mortality Data Ecosystem

DYLAN THOMAS DOYLE, University of Colorado Boulder, U.S.A
CASEY FIESLER, University of Colorado Boulder, U.S.A
JESSICA PATER, Parkview Health, U.S.A
JED R. BRUBAKER, University of Colorado Boulder, U.S.A

Access to mortality data is critical for researchers for various reasons, including the development of life-saving medical interventions. However, high-quality mortality data is typically unavailable for researchers to access in a timely manner due to bottlenecks and inefficiencies in the coordination of managing mortality data across a vast information ecosystem. To identify opportunities to improve access to mortality data, we conducted a qualitative interview study with 20 experts with direct knowledge and experience with mortality data systems. From these interviews, we mapped the mortality data ecosystem and found that policy constraints are the main underlying cause of bottlenecks and inefficiencies. We argue that policy intervention is required to optimize the coordination of mortality data between databases and between organizations. To support researcher access to mortality data, we present guiding principles for designers seeking to improve the mortality data ecosystem. These principles contribute to CSCW scholarship focused on the challenges of coordination across large information ecosystems and the tradeoffs between technology and policy when designing such systems.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing systems and tools**; **Collaborative and social computing theory, concepts and paradigms**; **Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: end-of-life support, data and accounts, policy, law, legislation, regulation CDC

ACM Reference Format:
Dylan Thomas Doyle, Casey Fiesler, Jessica Pater, and Jed R. Brubaker. 2024. Designing for Researcher Access in the U.S. Mortality Data Ecosystem. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 493 (November 2024), 31 pages. https://doi.org/10.1145/3687032

1 Introduction

The healthcare sector increasingly depends on big data systems to connect medical data to clinical researchers [5]. One area in which big data systems are progressively deployed is connecting clinical researchers to mortality data [1]. Traditionally, access to mortality data has been essential for researchers to inform public health strategies [95, 42] and to understand disease burdens [67], yet it is also becoming critical for the development of life-saving medical interventions [1, 29]. Successful oncology clinical trials, for example, depend on knowing the future health outcomes of their participants: those whose conditions improve, and those who die [29, 91]. If oncology researchers cannot get timely access to high-quality data about the death of participants, it is

---

Authors' Contact Information: Dylan Thomas Doyle, dylan.doyle@colorado.edu, University of Colorado Boulder, Boulder, Colorado, U.S.A; Casey Fiesler, University of Colorado Boulder, Boulder, Colorado, U.S.A; Jessica Pater, Parkview Health, Fort Wayne, Indiana, U.S.A; Jed R. Brubaker, jed.brubaker@colorado.edu, University of Colorado Boulder, Boulder, Colorado, U.S.A.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2024 Copyright held by the owner/author(s).
ACM 2573-0142/2024/11-ART493
https://doi.org/10.1145/3687032

difficult to develop the most effective medical interventions [54]. When high-quality mortality data is unavailable to researchers, the development of medical interventions can lag or, in certain circumstances, can be impossible to develop [21].

Several recent medical studies have found that high-quality mortality data is unavailable for researchers to access in a timely manner [21, 82]. These studies have focused on suggesting improvements to hospital death certificate charting practices. However, paying attention to charting practices alone will not adequately address data accessibility issues. After all, healthcare data flows through a wide array of vital records information systems from its creation in an Electronic Health Record to when a researcher can access it [22, 74, 23]. Focusing on individualistic improvements, like within hospital practices, is important but will only address a portion of the underlying issues that complicate mortality data access. To more comprehensively diagnose bottlenecks (situations where the flow of data is hindered, resulting in slower data processing and analytics) and inefficiencies constraining mortality data access, we must consider how mortality data is coordinated across a vast ecosystem of organizations and databases. We understand data ecosystems through the orientation of Ecological Systems Theory, which "posits that individuals constantly engage in transactions with other humans and with other systems in the environment, and that these individuals and systems reciprocally influence each other [27]." In HCI literature, ecosystem approaches informed by Ecological Systems Theory have been used to address issues of data access across vast networks of private and public information systems that have many stakeholders coordinating data movement [30]. Taking inspiration from these studies, we use an ecosystems approach to identify challenges to mortality data access that can account for the coordination of mortality data across a wide array of organizations and databases. Drawing from Star and Griesemer, we understand coordination here as the cooperation between the many different actors, viewpoints, and systems that manage, process, and exchange mortality data [93].

To identify opportunities to improve access to mortality data in the mortality data ecosystem, we conducted a qualitative interview study with 20 experts with direct knowledge and experience with mortality data systems. The experts we interviewed represented a wide array of organizations in the ecosystem: clinicians who create death certificate entries, funeral directors who amend those entries, state vital records office employees who archive and process those entries, federal government employees who interact with the National Death Index (NDI)[1], and clinical researchers. Based on these interviews, we mapped the mortality data ecosystem as seen in Figure 1. Our analysis and mapping identified bottlenecks and inefficiencies in the ecosystem that impeded coordination, leading to complications in researchers' ability to access data. These bottlenecks and inefficiencies had policy constraints as their main underlying cause. Given this ecosystem's decentralized and multi-organizational nature – and the policy constraints that we found as a main impediment to organizational coordination – we argue that technology solutions alone are inadequate. Rather, we posit that policy intervention is required to optimize the coordination of mortality data between databases and organizations, thereby reducing barriers to data access for clinical researchers.

To support researcher access to mortality data, this study presents guiding principles for future work in the mortality data ecosystem – informed by our empirical findings and prior CSCW literature. These principles contribute to CSCW scholarship focused on the challenges of coordination across large information ecosystems and the tradeoffs between technology and policy when designing such systems. In our discussion, we additionally consider how cultural relationships to death in the U.S. impact policy implementation. We conclude by discussing the impacts of data-sharing

---

[1] Containing over 100 million death records, the National Death Index (NDI) is the preeminent database for researchers to identify who in a study has died by linking research datasets to death certificate information for study subjects. The NDI matches clinical study subjects to U.S. death records and provides the date and causes of death for subject matches. Clinical researchers commonly use the NDI to calculate life expectancies and adverse risks from specific causes of death [22].

policies (e.g., HIPAA) on researcher access to mortality data and suggest paths forward for future work on this front.

2 Background

2.1 U.S. Vital Records Information Systems

We describe the mortality data ecosystem in more detail in Section 5. However, to clarify our study parameters, it is important at the outset to describe the roles and responsibilities of U.S. vital records information systems more generally.

Vital records information systems are the web of information systems and databases that archive and communicate health information [86]. In the U.S., vital records include births, deaths, marriages, divorces, and fetal deaths. They also record information about the cause of death [74]. At a federal level, records are kept by the National Vital Statistics System (NVSS) within the National Center for Health Statistics (NCHS) [75]. The NCHS is housed within the Center for Disease Control and Prevention (CDC), an operating division of the U.S. Department of Health and Human Services (HHS) [22]. Vital Records data are provided through contracts between NCHS and vital registration systems operated in the various jurisdictions legally responsible for registering vital events. In the case of mortality information systems, the number of reporting jurisdictions is 57, which includes all states, territories, and specially designated jurisdictions such as Washington, DC [73, 49].

Over 11 million vital events are reported annually in the United States, including 4.3 million births, 2.4 million deaths, 26,000 fetal deaths (over 20 weeks gestation), 2.1 million marriages, 1.2 million induced terminations of pregnancy are reported, and 1.1 million divorces [86, 36, 65]. A report by Shwartz [86] summarizes the complex archiving and processing of these records with the following:

> Depending on jurisdiction laws... events are reported directly to state and local vital statistics offices in the form of certificates of birth, death, marriage, and divorce and reports of spontaneous and induced termination of pregnancy... State and local vital records offices, as the custodians of the records, must register and process these original, generally confidential records. Jurisdiction systems vary widely and by certificate type, which ranges from certificates that may be printed on 100 percent cotton paper with wet signatures for archiving to state-of-the art paperless records received through a secure Internet transmission with a biometric authentication. Regardless of the method, each record is unique, must be logged in, evaluated for data quality, and assigned file numbers.

The vital records have both legal and public health purposes [86]. This study is interested in the public health purposes of one type of vital records; mortality records. Local (e.g., county-level) and state vital records offices are responsible for death certificates and safeguarding the information they contain (fact of death, date of death, cause of death, for example). Again summarized by Schwartz:

> Intrinsically, local or state governmental function is both a drain on vital statistics offices' human resources and a source of revenue for jurisdictions. The revenue is substantial. Birth certificate fees among states in 2007 ranged from 9 (Florida) to 30.00 (New York), with an average price of about 15.00 per certified copy. Collectively, jurisdictions issue millions of certified copies of birth and death certificates annually. Unfortunately, the revenue often goes to a state's general revenue fund and not back to the program. Thus, personnel are often diverted from other work to meet important—and highly visible—customer service needs. As a result, less visible functions, such as quality assurance, training, and hospital site visits, may receive less attention or

> fewer resources... Every jurisdiction is responsible for the receipt, processing, quality, and analysis of its data. Vital-event data are derived from the certificates filed with each jurisdiction and become the vital statistics of that jurisdiction.

Relative to the *identified* mortality data this study is interested in, U.S. vital records information systems are the series of databases and organizations (from the hospital level to the local level to the state level to the federal level) that archive and process death certificate data on its way to being accessed by clinical researchers. We describe the type of vital record information we are interested in, identified mortality data, in the following section.

2.2 Types of Mortality Data and Study Scope

Mortality data can be either de-identified or identified – typically classified as fact-of-death data (de-identified) or cause-of-death data (identified) [53]. The differentiating factor is that identifiable cause-of-death data is linked to the name and social security number of the deceased [48]. In this study, we focus on identified mortality data, as it has a greater utility to clinical researchers [57]. For clarity of our research focus, it is worthwhile to distinguish the two types of data.

Both de-identified and identified data are useful for researchers, albeit in distinct ways. Broadly, de-identified data is useful for mapping statistical trends [57], and identified cause-of-death data is useful for developing clinical interventions to respond to those trends [35]. De-identified data is useful for researchers conducting statistical analyses, such as those interested in nationwide trends in death rates [88]. Providing an option to obtain de-identified data is important to protect privacy since individual death records cannot be tracked to the deceased's identity. De-identified statistical data may also provide insight into current trends and predict future trends in global and federal mortality factors [97]. For example, during the COVID-19 pandemic, epidemiologists used CDC mortality data to predict spikes and troughs in infection rates [81]. Data to map high-level trends do not need to be identifiable.

Identified cause-of-death data has more utility to clinical researchers to track health outcomes for specific individuals and groups who have died following a study or a certain treatment [28]. For example, a researcher at a medical school testing a new drug for an oncology intervention needs to access information about the death of a former patient who has moved out of state and has died. Identified data is much more valuable for clinical researchers designing new drugs or other patient-facing interventions (e.g., an oncologist conducting clinical drug trials wanting to follow up on when/how a person previously undergoing their trial died) [83]. De-identified data allows for broad-sweeping statistics that protect privacy after death, but it does not allow for deeper context, including individual or group follow-up to assess interventions over time – the main use value of mortality data for clinical researchers conducting patient-based studies.

3 Related Work

To motivate and provide context to our work, we turn to prior research on U.S. vital records information systems, post-mortem data management, and policy and information systems.

3.1 Challenges for U.S. Vital Records Information Systems

Research on the challenges that U.S. vital records information systems encounter has largely been published in medical journals and has focused on improving the Electronic Health Record (EHR) at the hospital level [87, 71, 61]. Work has primarily been conducted on EHR data quality assessments and EHR informatics, suggesting practices like standardized documents that can improve data collection and utility [13]. This prior work has identified data quality issues as the primary challenge to improving the utility of health data for clinical researchers and has focused on designing new standardized versions of death certificates to solve those challenges.

HCI research, meanwhile, has primarily focused on vital records information systems at a higher level than medical research. HCI researchers have been interested in developing healthcare interventions in big data systems [102, 24] and health data confidentiality [9]. Outside of vital records, other recently growing areas of healthcare research are personal health informatics [70], self-tracked data [51], and building big data healthcare platforms [64]. In the past few years, HCI has expanded this work on vital records by taking an interest in organizational improvements for health records – particularly through improving documentation practices [78].

Regarding mortality data in vital records information systems, previous medical research has predominantly focused on designing interventions to improve the accuracy of death data [41]. Literature has especially focused on designing more effective checklists for clinicians filling out death certificates, and streamlining physician training on EHR reporting [15, 47, 57]. Research has found that without high-quality data, there are dangers of clinical researchers generating incorrect models or misidentifying patient outcomes [57, 21]. Scholarship has considered high-quality death data as data that capture identifiable information about the decedent and delineates the primary and underlying causes of death [57, 21].

HCI research on mortality data managed by vital records information systems has been scant. The few studies that have been conducted have focused on improving the validity of cause-of-death on death certificates [52] (similar to medical literature), and applying NLP models to optimize death certificate topic modeling [55]. To our knowledge, there has been no attempt in HCI literature to engage with mortality data at a federal level – or to consider researcher access of mortality data through the NDI. This blind spot is concerning, considering that medical researchers access mortality data almost entirely through the NDI [22, 73].

Responding to the lack of a cohesive data ecosystem map, this study maps the mortality data ecosystem of identified cause-of-death data within U.S. vital records information systems. This study expands medical literature on improving data quality by considering challenges to data quality beyond the EHR. We additionally complement HCI research on healthcare and big data by identifying bottlenecks and inefficiencies throughout a large-scale mortality data ecosystem – and beginning to develop potential solutions to address them.

3.2 Post-Mortem Data Management

In contrast to limited research related to mortality data in vital records information systems, there is substantial CSCW scholarship on the post-mortem management of personal data on accounts and platforms [18, 19, 46, 68, 76, 66].

We know from previous literature that a social-technical gap exists between the needs of dying and bereaved users and the technical capacity of information systems to fulfill data management needs [32, 34, 76, 66]. For example, Gach and Brubaker discuss a gap between the social need of trust when designing online memorials, and the technical capability of social media to account for that trust [43].

Post-mortem data management is clearly of interest to CSCW. However, post-mortem data encompass more than the data and accounts of deceased individuals – it additionally includes health records and vital records generated in the wake of someone's death. This study expands CSCW's understanding of social-technical gaps in mortality data by looking at a different type of data than has been previously studied. Previous research has studied personal-facing data generated by a user that must now be managed post-mortem, and has largely focusing on social media platforms [18, 19, 16, 17, 43, 44, 33, 31]. This study instead examines mortality data captured by U.S. vital records information systems. By expanding the type of death data being studied, we can begin to see similarities and differences in the needs of death data management across different kinds of information systems.

3.3 Policy and Information Systems

As we will describe in Section 6, our findings indicate that policy barriers are the main underlying cause leading to bottlenecks and inefficiencies in the mortality data ecosystem. Because policy can have multiple definitions and understandings [6], it is important to clarify what we mean by policy. In this work, we understand policy per Jackson et al.'s definition of the term [56]:

> Policy includes the forms of public law that regulate technology design and use... the public bodies and processes charged with establishing laws and regulations... and the public agencies charged with implementing them: the courts, state regulators, and a wide variety of administrative agencies... [Policy] also includes a wide range of rules and procedures set by private firms – software licenses, end user license agreements, policies for app developers, and at the limit corporate philosophies (where actually backed and followed through in organizational practice) – that may intersect with formal law and policy in complex ways.

Early HCI and CSCW scholarship strongly considered the role of policy and its governance within organizations. For example, the impact of company policy on groupware integration in an organization was identified as a key factor in the adoption of groupware within that organization [45]. In recent years, scholarship in this area has focused on aspects such as copyright [39], AI regulation frameworks [20], and worker-centered design [108], to name a few. Especially relevant for our study, in the past decade CSCW has shown increasing interest in how policy impacts data coordination in large-scale healthcare systems and scientific research systems – with a focus on the impact of public policy on practice [37, 90, 77, 38]. This previous work has demonstrated that the impact of policy on the coordination of data in healthcare and research settings is of importance to social computing scholars in CSCW and CHI communities. Our work contributes to this growing body of policy-based research by examining how public policy impacts the practice of mortality data coordination in healthcare, research, and government systems.

Although the above recent research has demonstrated that policy is a key mediator for social-technical gaps, in the overall area of CSCW research, policy has often been an afterthought compared to technical practice and design [56]. In surfacing the importance of policy interdependencies in their articulation of the policy knot model, Jackson et al. summarized that "policy is tucked away in the future research section of a paper focused on some other research area [56]." Our work provides a further case study of the importance for CSCW to center policy in its scholarship.

Bringing policy into the forefront of our analysis, in our Discussion we use Jackson et al.'s Policy Knot model to interpret our findings. In this study, we draw on the policy knot model as an analytical lens to interpreting data ecosystems through examining interdependencies between policy, design, and practice – arguing that by focusing first on the interdependencies between policy, design, and practice HCI researchers and practitioners can more effectively identify areas of challenge and intervention into a data ecosystem.

Previous HCI research has used the policy knot model to consider futures for HCI-policy collaboration [107, 92], the dangers of techno-policy solutionism in child safety online [4], data privacy concerns in globalized markets [100], and determining accountability for traceable data sets across organizations [60]. In this work, researchers have used the policy knot as an analytical lens to determine tensions in data coordination between stakeholders (e.g., local organizations vs. national organizations vs. international organizations [103]) and to suggest new perspectives for HCI researchers and practitioners to apply as they develop solutions to data coordination challenges (e.g., bringing attention to the pitfalls of a techno-policy solutionist mentality [4]). Taking inspiration from this work, in this study we apply the policy knot as an analytical lens to determine tensions in data coordination and identify principles to guide future intervention. We expand previous work that has used the policy knot model as an analytic lens by providing an illustrative case study identifying policy interdependencies for the coordination of mortality data across the healthcare and government sectors.

4 Methods

4.1 Recruitment

We conducted a qualitative interview study that included 20 semistructured interviews with experts on U.S. mortality information systems. We considered an expert to be someone who had interfaced with mortality information systems in their professional life for more than 5 years. We sought to interview experts from a variety of perspectives and a diversity of ecosystem domains. Specifically, we sought to interview experts in the medical sector, the funerary sector, the state government sector, the federal government sector, and the clinical research sector. We identified these groups because they represent each step of the mortality data flow – from death certificate creation to researcher access (see Figure 1). Leveraging the professional network of the research team [specifics omitted for blind review], we recruited a convenience sample of 3 participants from the clinical research sector for preliminary interviews.

We used snowball sampling from these first 3 researchers, receiving 8 recommendations for participants in the funerary, state government, federal government, and clinical research sectors. After conversations with the next 8 experts, we used snowball sampling to recruit an additional 9 participants. In all, experts represented the following areas: medical (n=4), clinical research (n=4), federal government (n=5), state government (n=4), and funerary sectors (e.g., funeral directors and morticians; n=3). State government representatives were from the states of California, Colorado, and Oregon. Clinical research participants were from R1 medical research academic institutions in the states of California, Colorado, and New York. Participants were approached in the capacity of experts, but not in the capacity of a formal representative of their organizations. Participants spoke from their own experiences and their views did not necessarily represent the views of their employers. We summarize our participants in Table 1.

We chose to utilize snowball sampling since it allows for leveraging specialized networks [72], which is needed to find experts in niche areas such as mortality information systems. A limitation of snowball sampling is that there is a chance it creates an echo chamber of participant perspectives. Although we strategically sampled through a theoretic sampling of the most important stakeholders in mortality information systems, future studies should consider widening the scope of stakeholder groups to increase the generalizability of our findings. Additionally, this study is limited by our exclusive focus on mortality data in U.S.-specific vital records information systems. As we expand on in the Limitations section at the end of this paper, this study is not necessarily generalizable to other types of data in vital records information systems (e.g., birth certificates), and is not generalizable to other geographies.

4.2 Data Collection

Interviews ranged from 45 to 90 minutes and were conducted remotely via Zoom. All interviews were recorded. The data collected included interview audio recordings, transcripts, and fieldnotes. The interviews consisted of two parts, described below.

Participant	Gender	Area
P01	F	Medical
P02	M	Medical
P03	M	Medical
P04	F	Medical
P05	M	Clinical Research
P06	F	Clinical Research
P07	F	Clinical Research
P08	M	Clinical Research
P09	F	Federal Government
P10	M	Federal Government
P11	F	Federal Government
P12	F	Federal Government
P13	F	Federal Government
P14	F	State Government
P15	M	State Government
P16	M	State Government
P17	M	State Government
P18	M	Funerary
P19	M	Funerary
P20	M	Funerary

Table 1. A table of all participants, including their gender and professional area of expertise.

4.2.1 Interview Part 1: Data Flowchart Feedback and Amendment. In the first part of the interviews – to better understand the data ecosystem – we asked participants to provide feedback on an initial flowchart representing the mortality data ecosystem.

The first goal of this study was to map a mortality data ecosystem, since, without such a map, it would be difficult to contextualize ecosystem challenges [8]. To our knowledge, no previous map existed that mapped the data flow of mortality data across organizations and stakeholders. Although public-facing federal documents such as [22, 23] outline the complex organizational hierarchy of who reports to whom, these documents do not clearly describe how data moves between the databases in these organizations. Therefore, prior to the interviews, the research team had created an initial visualization of the data flow of death certificate data from the hospital to the NDI-level. When constructing our initial flowchart, we sought to trace the data flow of mortality data from the point of its creation in the death certificate following a death event to the point of researcher access from the NDI. Because our focus was on the data ecosystem, in our analysis we paid special attention to the transitions of data between organizational databases (e.g., when data is transferred from a state database to a federal database).

To visualize this initial data ecosystem – visualized as a flowchart – we first turned to relevant academic literature and descriptions of the U.S. vital records information systems referenced in Section 2 [86, 73, 23, 74, 22]. Identifying areas in the literature where it discussed moments where data was transferred from one organization to another, we generated the initial flowchart depicting those transfer points in Figma.

Having established an initial flowchart, in our interviews, we reviewed an evolving version of the flowchart with participants to amend and corroborate it. Specifically, we asked participants to elaborate on the evolving flowchart and provide feedback based on their professional area of expertise, focusing on how data moves through the ecosystem (e.g., what data is moving, through which databases, in which instances does data move, etc.). After all interviews were conducted and we had a well-corroborated version of the flow chart, we requested final feedback from participants and then created the final visualization seen in Figure 1.

4.2.2 Interview Part 2: Identifying Bottlenecks and Inefficiencies in Data Flow. In the second part of the interviews – to examine the bottlenecks and inefficiencies within the mortality data ecosystem – we asked participants to discuss the data transitions in the area of the ecosystem they are professionally located in. For example, participants located in federal organizations were asked specifically about how databases like the NDI receive, process, and share data.

4.3 Data Analysis

Following data collection, data analysis occurred in two stages: (1) analysis to map the data ecosystem, and (2) analysis to inductively identify themes from interviews. For the first stage – mapping – coding was deductive, with researchers incorporating participant feedback on the flowchart to validate and amend our understanding of the data ecosystem. For the second stage, we used Braun and Clarke's process for thematic analysis [14], informed by Saldana's approach [84]. Our thematic analysis focused on identifying bottlenecks and inefficiencies, informed by our desire to understand the challenges of high-quality mortality data being unavailable for researchers to access in a timely manner. Guided by our focus on identifying bottlenecks and inefficiencies, coding was inductive (iteratively and reflexively grouping codes into similar categories).

During both stages of data analysis, members of the research team independently listened through each of the recordings and reviewed field notes, thereby familiarizing us with the data. We then read through the transcripts, employing closed coding for stage one to amend the flowchart – following the process described in 4.2.1.

After the flowchart was finalized, researchers began stage two of analysis, again familiarizing ourselves with the data. Interview transcripts and field notes were uploaded as Microsoft Word documents into a secure Microsoft OneDrive folder and organized by participant number for coding purposes. Each research team member had their own copy of each interview to annotate. Researchers independently annotated each interview, open coding with a focus on identifying areas of bottlenecks and inefficiencies described in the interviews – the focus of our research question. After independent open coding, researchers came together to construct a codebook in a shared Microsoft Excel spreadsheet, generating initial codes by sharing the results of their open coding and discussing them as a group.

From our discussions, we generated an initial codebook. For example, initial codes included *costly NDI access fees*, *slow matching of identity to state records*, and *incomplete NDI spreadsheets.* The codes in this initial codebook were typically illustrative of specific challenges participants had mentioned.

After a preliminary codebook of these more specific challenges was generated and populated with initial codes, we looked for relationships between those codes to construct preliminary themes. Theme memos were written for preliminary themes [14]. The research team then considered how the preliminary themes in the theme memos related to one another, generating overarching themes. For example, grouping the *costly NDI access fees* code with codes such as *cost differentials between organizations*, and *cost of individual death certificates* created a preliminary theme of *cost.* When reflecting on how the different aspects of cost related to one another – and related to our interest in bottlenecks and inefficiencies – we decided on a final theme category of *Prohibitive Cost of Data.* We describe each of these final themes, which focus on symptoms and underlying causes of bottlenecks and inefficiencies in the data ecosystem, in Section 6.

Throughout the above coding process, the research team iteratively returned to the data. As an inductive thematic analysis process, themes were written, edited, and reorganized during our discussions in each coding step. Before moving on to a new step in the coding process, all research team members reached a consensus on codes and themes. In the case of disagreements, researchers discussed, noted differences, and continued the discussion until a consensus was reached. Researchers sought thematic saturation. If thematic saturation was not achieved after stage two analysis, additional recruitment, interviews, and data analysis would occur. The research team reached a consensus that thematic saturation had occurred, and so no new participants were recruited.

We present the results of our analysis in two sections. First, we present the findings of our mapping of the mortality data ecosystem, synthesized from public-facing organizational literature (e.g., CDC information pages, State Vital Statistics reports, legal statutes, etc.) and corroborated by participants in interviews. We then present the thematic findings of our interviews – describing the symptoms and underlying causes of bottlenecks and inefficiencies in the mortality data ecosystem.

5 Findings: Mapping the Data Ecosystem

The ecosystem flow chart we developed (see Figure 1) summarizes the flow of mortality data at a high level – highlighting four main phases of data transition in the database flow column. This ecosystem map is current as of October 2023.

We note that organizational policies, regulations, and laws can change quickly depending on presidential administration and any number of other political factors. We also note that a possible limitation of this flowchart is that areas of the data ecosystem may not be accounted for in the academic literature, public-facing government documents, and discussions with the 20 experts we consulted – and therefore, those areas would not be represented in the chart. Additionally, because it is such a large data ecosystem, some nuances and blindspots are difficult to capture in such a high-level chart. Therefore, we present this flowchart as a dynamic reference point to begin to understand this data ecosystem and not as a comprehensive representation of every organization and facet. We sincerely hope that future research builds on this initial flowchart to bring greater attention to the nuances and blindspots we may have been unable to capture.

5.1 Ecosystem Overview

Including the NCHS system, data about people's deaths in the U.S. are stored, administered, and accessed across databases managed across four sectors: the medical sector (e.g., the Electronic Health Record/EHR), the state government sector (e.g., the Electronic Death Record/EDR), the federal government sector (e.g., the NDI), and the clinical research sector (e.g., academic institution databases). Much more recently, the private sector (e.g., pharmaceutical companies) increasingly plays a role in the brokering of mortality data [94]. However, this role is currently so small and new that it is difficult to account for, so we do not include it in our visualization. Future research should consider how the private sector is impacting mortality data access.

We include both county and state government organizations in the state government sector of the flowchart. Although, in some states, the county level is a distinct database (e.g., Indiana), in other states, county vital records offices have been closed in favor of more centralized state-level offices (e.g., California). In other states still, county vital records offices use the same database system as the state vital records office (e.g., Massachusetts). For clarity and consistency with what experts highlighted in our interviews (e.g., interviewees rarely spoke of the county level and almost

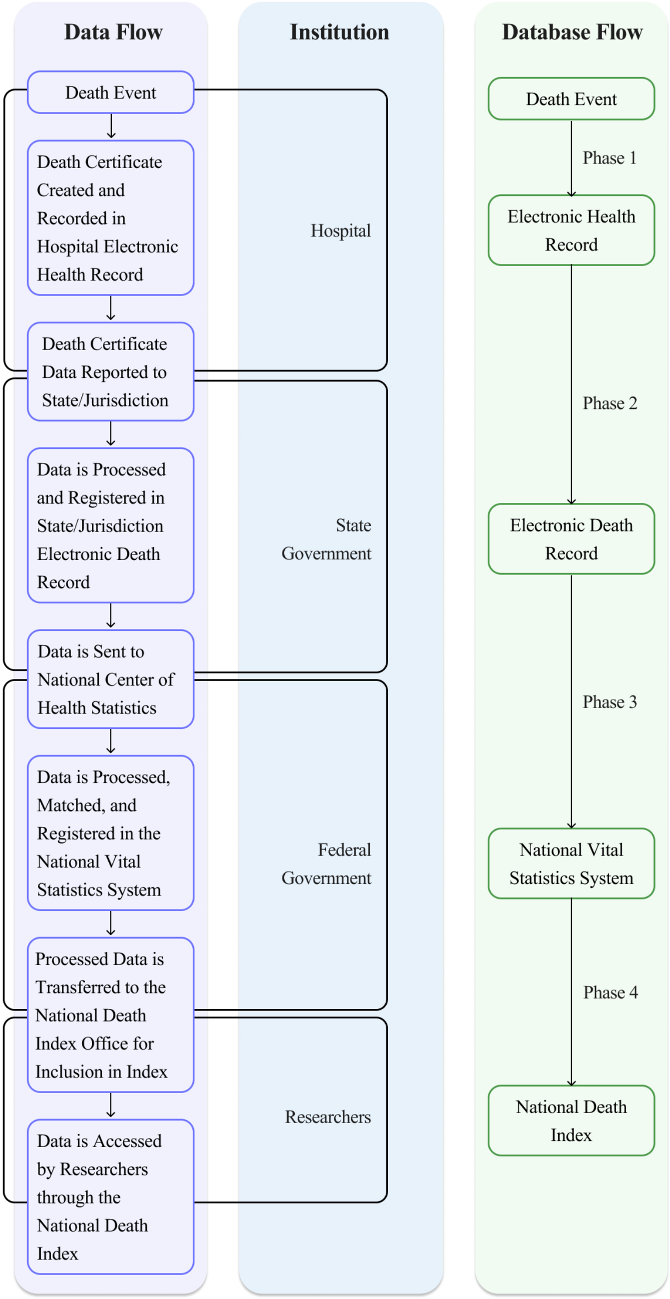

Fig. 1. A visualization of the mortality data ecosystem. This flowchart depicts three columns: (1) the data flow of mortality data from the moment of creation in the death certificate to access in the NDI, (2) the institutions that manage each step in the data flow, and (3) the database flow as data moves through the ecosystem. Boxes around institutions and parts of the data flow overlap to signify when multiple institutions are involved with the storage or processing of data as it moves between databases. The database flow column describes four main phases of data transmission as mortality data moves between databases in institutions.

always lumped the county and state level as the same phase), we have collapsed the two into a single phase for the purposes of visualization and summary.

5.2 Mortality Data Flow

We summarize the mortality data flow in U.S. vital records information systems in Figure 1. This figure visualizes three interrelated flows that make up the data ecosystem. First, this figure visualizes the flow of mortality data from the point of the death event in a hospital/health system to the point that the data is accessed by researchers through the NDI. Second, this figure visualizes the institutional flow as data moves from the hospital to the state government to the federal government to researchers. Third, and most importantly for our findings about bottlenecks and inefficiencies in the ecosystem, this figure visualizes four phases of database flow – as mortality data moves from the death event to the NDI. We describe these phases in more detail below.

Phase 1: Data Creation. In the U.S. context, data about people's deaths are typically generated at the hospital level when a physician creates a death certificate at the time of a person's death[2]. The death certificate record is created in the Electronic Health Record (EHR) in a hospital's charting database.

Phase 2: Electronic Health Record to Electronic Death Record. Following the creation of the death certificate, hospitals submit death data to state-level vital records jurisdictions so that the vital event can be reviewed, edited, processed, and officially registered. In some states, data is automatically shared with the Electronic Death Record (EDR) belonging to the jurisdiction where it was generated without the hospital system having to upload it manually (e.g., data generated by a California death certificate is automatically ported to the EDR managed by the Vital Records Office/VRO of California). Depending on the jurisdiction, in some instances mortality data passes through a county-level VRO before being sent to a state-level VRO.

The data fields typically included in an EDR record are personal identifiers (e.g., SSN, age, name), demographic information (e.g., race, ethnicity, gender), and cause-of-death information (e.g., the primary and secondary cause of death as determined by the clinician generating the record). In some cases, the VRO may undertake further quality control and reach back out to data providers to confirm that the data file contains correct information. The VRO in the state jurisdiction certifies the death certificate, checking the identity and certifying the cause of death of the deceased. Following certification and registration, death certificates can be bought from the VRO by family members of the deceased.

Phase 3: Electronic Death Record to National Vital Statistics System. The CDC, via the NCHS office, then receives and processes the data in its own database (the National Vital Statistics System). State VROs report mortality data annually – after finishing year-end edits and processing (e.g., data cleaning). The state VRO submits the final data files to NCHS at the annually appointed time. Though the CDC receives mortality data from every jurisdiction, laws dictate that the mortality data is legally owned and administered on a jurisdiction level (e.g., Oregon owns and administers mortality data from Oregon, Nebraska owns and administers mortality data from Nebraska, and so on). A third-party non-profit, NAPHSIS (National Association for Public Health Statistics and Information Systems), brokers the financial contracts that 'rent' mortality data from the state jurisdiction to the NCHS.

---

[2] Death certificate data in the U.S. is almost always generated at the hospital level. However, exceptions exist, such as when a police coroner conducts a lengthy investigation into a mysterious cause of death or suspected foul play. For clarity, this study focuses on the most common cause of mortality data creation: death certificate creation in the hospital setting

Phase 4: National Vital Statistics System to National Death Index (leading to public statistics reports and researcher access of data). After a final round of processing, matching, and cleaning, the NCHS makes data available to the public and, upon special application, to clinical researcher groups. The NCHS produces two access points to data: public statistical reports (public reports with de-identified data), and the NDI (private reports with identifiable cause of death data, only available upon application and vetting). The purpose of the NDI is to provide researchers access to identifiable data (data that matches the decedent's identity with their cause of death, demographics, and other contextual data points such as their time of death).

5.3 Summarizing Data Flow

We identified four phases of data transition between organizational databases, marked by moments when data moves from one organization's database to another. Each transition is unique regarding stakeholders, data processing policies, and specific database software utilized. A limitation of this kind of ecosystem mapping is that the data lifecycle can appear universal, whereas we will see below (see Section 6) that data transitions can be heavily influenced by variable local policy. However, this ecosystem map is helpful in providing a common understanding of the data flow in broad strokes so as to better understand its bottlenecks and inefficiencies.

6 Findings: Underlying Causes of Bottlenecks and Inefficiencies

Through our analysis, we found that researcher access to mortality data is constrained by bottlenecks and inefficiencies throughout the entire data ecosystem. To clinical researchers seeking to access mortality data, we found that the symptoms of these bottlenecks and inefficiencies are: (1) prohibitive cost, (2) lack of timeliness, (3) archaic government data systems, and (4) lack of data quality (lack of consistent, complete, and accurate data). We found that the underlying causes of these symptoms are mainly a function of (1) incoherent policy, and (2) cultural aversion to engaging with death.

The issues of prohibitive cost, lack of timeliness, archaic technical systems, and data quality are not particularly novel when it comes to identifying issues with healthcare data [102]. What is novel is our finding that these issues are, in fact, emergent symptoms of underlying challenges – incoherent policy and cultural aversion to engaging with death.

6.1 Impacts of Policy

Through our analysis, we found that incoherent policy is the main underlying cause of bottlenecks and inefficiencies in the mortality data ecosystem – leading to pernicious challenges to the coordination of data between organizations. By incoherent policy, we mean policies (e.g., rules, regulations, and laws [56]) across different areas of the data ecosystem that contradict or complicate one another, leading to lag or other complications in the technical flow of data between databases. We found prohibitive cost, lack of timeliness, archaic technical systems, and lack of data quality to be the main symptoms of poor coordination in the data ecosystem – but the underlying cause was problems with policy. In this section, we describe the symptoms and their relationship with policy complications.

6.1.1 Prohibitive Cost of Data. The prohibitively high cost of mortality data was the most commonly referenced barrier to researchers accessing mortality data. Experts indicated that the rules and procedures enshrined in government policy in some cases – and the absence of regulatory policies in other cases – facilitated the sustained high costs.

Experts reported that researchers must pay prohibitive fees to access NDI records. Each decedent's record costs the amount that the state jurisdiction has determined it to cost. NCHS policies dictate that a record costs the same price point a layperson would pay. State jurisdictions charge the NCHS similar prices per death certificate. For example, if a Colorado resident has to pay 25 dollars for a death certificate, the NCHS must pay the same 25 dollars per record. This cost structure, determined by state law, cascades to research groups, who must pay the same 25 dollars to the NCHS. Upon receipt of payment, the NCHS then brokers a contract with the research institution and provides a file of a number of identified death records that the research institution has access to and can use per NDI rules.

One interviewee summarized the cost challenge as follows:

> If you got the price right, every hospital in the country would buy that dataset. As it is now, the data is not affordable to researchers. We can't do our work without the data, but there is no way an academic research lab can pay that kind of asking price. (P03)

Interviewees reported that the high cost is due to unsustainable cost structures in state and federal budgets, resulting in death certificate sales being a significant revenue stream for state vital records offices. Vital records offices and state governments keep the price high out of necessity to keep vital records offices functioning. In other words, the high cost is due to a limited budget for state vital records offices, determined by state and federal budget policy.

P07 noted that there is a cyclical relationship in the unsustainable cost structure: the federal government assumes that the state government will financially support death data management, the state vital records offices assume that the state government will support death data management, and the state government assumes that the vital records offices will financially support death data management through its death certificate revenue stream. Interviewees reported that the financial system is so entrenched with assumptions about who has the responsibility to pay for the state-level offices that manage mortality data, enshrined in state and federal policy, that there is no easy path forward without disrupting political dynamics:

> In many states, the vital statistics office that [handles death data] is not funded by the state budget. Even in cases where it is, the assumption is that [the state vital records office] is self-sufficient, since they are one of the only offices that has a product they sell. Little pieces of income come in through contracts, but it's not enough. The offices often scrape by relying on people buying certificates... All of these budget policies lead to [researchers] not being able to get what they need because the inefficient system inflates the cost. (P12)

A further factor contributing to prohibitive costs is the consequence of competing entities that can pay for the high cost. Competing entities such as the Social Security Administration can purchase mortality data from states using a much higher budget than research groups typically have access to. Simply, large government entities with large allocated budgets can pay the cost, whereas clinical research groups cannot. Although research groups are important beneficiaries of mortality data access, they are less lucrative consumers than large government entities purchasing the data. Since the state can make relatively substantial revenue selling mortality data at a high cost, they have no incentive to limit the cost for smaller consumers, such as researchers accessing mortality data through the NDI. Summarized by P03, "The SSA pays 9 million for the death data each year. Researchers can't keep up with that."

Experts noted that the unpredictability of the cost of data caused by incoherent policy exacerbated the logistical and administrative overhead for access. Participants said that the largest issue relative to unpredictability was that there is no law dictating a centralized pay scale to coordinate budgets or expectations around – each state controls its own pay scale. P02 summarized the overall challenges of mortality data cost as follows:

> How do we improve the way that analysts are able to get the data more easily? How do you liberate that data? Consider the pricing and cost structure. There is a perversion

> of paying for death certificate data in the first place. And then add on top of that that each state has its own rules to play by. And then add on top of that that you never know who you should be talking to, and you have a very messy system that's hard to work with.

6.1.2 *Lack of Timeliness (Lag in Data Reporting, Data Processing, and Researcher Approval).*

Researchers shared that more timely data access means more timely data analysis and that timely data analysis would mean more timely clinical solutions. State and federal government employees agreed that timely clinical solutions would mean more timely public health interventions. However, participants cited lag in data reporting, data processing, and researcher approval as a core challenge to mortality data access. Participants attributed lag to a lack of uniform regulatory policies across different organizations.

Participants considered timeliness as the amount of time it takes for mortality data to move through the pipeline: from data capture (e.g., the hospital), to state certification (e.g., the State vital records office), to data cleaning (e.g., the State vital records office, NDI), to the researcher application process (e.g., the NDI), to the actual reception of datasheets by researchers (e.g., the NDI). Participants stressed that the largest barrier to timely access was lag specifically as data is transferred between organizations and offices.

For example, when data is transmitted from the state vital records office to the NCHS and NDI office through the sharing of spreadsheets, there are often stumbling blocks for transferring it seamlessly. In this example, one stumbling block is the NDI office needing to cobble together differently-formatted data spreadsheets across the different reporting jurisdictions into a single file. Putting together that spreadsheet out of disparate sources takes time, energy, and employee labor, which invariably limits data access's timeliness downstream. The differently-formatted spreadsheets between organizations are a function of internal rules and policies within organizations. These rules and policies are not standardized across organizations, leading to lag in spreadsheet processing and, concordantly, downstream lag in researcher access to the data.

Lag in data reporting. Interviewees highlighted that one barrier to timely data access is the lag of data reporting at the state level. Data gets moved between the hospital, the county, and different state offices. Eventually, the state vital records office reports its mortality data to the NCHS. When the state reporting of mortality data to the federal level is not timely, the NCHS cannot process data promptly. The NCHS must process data in a timely manner to make data available for researchers to access through the NDI in a timely manner. The complexity of data reporting challenges was summarized by one participant, reflecting on data reporting lag during COVID-19:

> Well, we've got like weekly data on COVID-19 that we're pushing out on a regular basis. We've some monthly drug overdose information, although drug overdoses typically have a much longer lag, because the death investigation typically needs to be done in those cases, and the death investigation could take a few months, sometimes as many as 6 months. So, the cause of death information comes in later, even though we get the fact of death information in a very timely fashion. The fact of death comes in with the cause of death pending investigation... And then, once the death investigation is done, the death certificate gets amended with the new cause of death information, and then that information comes to us and we update our record. (P07)

This participant highlights one important complexity of reporting mortality data: the data being reported is not stagnant. Rather, mortality data records evolve over time – being amended across multiple databases simultaneously by multiple stakeholders. In P07's case, COVID-19 death data needed to be processed and certified quickly, but the hospital clinicians and coroners conducting a death investigation could take quite a long time to close a case. This participant also highlighted the complex relationship between fact-of-death data and cause-of-death data in databases, two interrelated types of data that get reported with different timeliness and sometimes have to be condensed into a single record. Hospitals, state reporting, and federal reporting have different regulatory standards for the ongoing updating of records, which gets increasingly complicated by instances of sudden death or deaths where foul play is suspected.

Beyond the hospital context, the difficulties of data reporting become even more complex when looking at the movement of data from the state jurisdiction level to the federal level. For example, timely public reporting for nationwide data lags due to legal precedent dictating that NCHS reporting (including the NDI) must abide by the reporting of the slowest state. In other words, the NCHS can only release annual reports on nationwide data once all states have reported their data. According to participants, the lag in reporting from the state is due to regulatory bureaucracy limiting the speed at which data can be certified (e.g., multiple levels of management must sign off on the data certification). Meanwhile, participants described that the reporting lag at the federal level is due to laws restricting data availability until every jurisdiction is accounted for.

To illustrate the messiness of annual reporting realities, one participant shared:

> For annual federal data, the release occurs about 11-12 months from the end of the day to year... Obviously, everyone would like it to be sooner. But at this point, that's about as fast as [the NCHS] can do it. Given the timing of the information coming from the states... [the NCHS] has all of the records by September or so, and then it takes a few months to prepare the data for release. (P17)

P08 provided an illustrative example of the degree to which delayed state reporting can impact the timeliness of the entire ecosystem. P08 cited one instance when the reporting lag was particularly poor – where it took 5 long years for the NCHS to finally receive the last bit of information out of one state. Due to legal regulations, it was only then that the NCHS could release that data as a federal report.

Lag in data processing. The slow speed of data processing (e.g., matching, linking) at the NDI level also causes researcher access lag. Participants shared that the slow speed of matching and linking is due to the thoroughness that the NDI is required by law to conduct their task. Participants who have previously conducted matching protocols were well aware of the importance of timeliness for research studies – with one participant even acknowledging the speed of academic publishing noting that once the data gets out of data, journals do not want to publish it. Both government and researcher interviewees highlighted the importance of timely matching, but also highlighted that even if they desired to change the system, they would be unable to without regulatory intervention from Congress.

Lag in researcher approval. The lengthy approval process that research groups must undergo before being given access to mortality data causes additional lag. The approval process is an outcome of rules determined in-house via the NCHS. As one researcher summarized:

> [The NDI] goes through a list of criteria that people have to read before they even apply... For some groups it's not worth their effort to apply because it takes time. The application process is typically anywhere from 6 weeks to 3 months. There are some applications that take up almost a year... And if groups are asked for revisions, if [the NDI] doesn't get your revision soon, they close out your application. But there is no way of knowing when your application will be processed and so it is hard to know when your response will be needed. (P10)

Participants reported that the NDI application has had to become stricter in recent years because the number of applicants has grown significantly in response to research in the aftermath of COVID-19. Applicants must demonstrate that their research is on public health and medical research. They must also demonstrate that they can safeguard the confidentiality of the data and limit access to only those people approved in the application.

The NDI office manually reviews applications. If a research group cannot adequately demonstrate the above in their initial application, the application is sent to an independent board of advisors. Whether an application is referred to the board is at the discretion of the NDI office at the CDC. The board is not affiliated with a branch or an organization within NCHS. Board members are a group of typically about 20 people from different parts of the U.S. – including representatives from academia, private industry, and federal, state, and local agencies. The board of advisors also ensures that applications adhere to state jurisdiction requirements. Participants said this thorough approval system was due to a fear of potential abuses of using mortality data. For example, one participant cited identity theft and targeted harassment of bereaved individuals as potentially harmful outcomes if the privacy of a decedent's name and cause of death were released.

Participants reported that the process is thorough but slow. Some researchers suggested alternatives that would speed up the current process. For example, one participant shared:

> Timeliness might be fixed if we move to a system that is preliminarily self-attest – if you can say that you meet these very specific criteria and are able to prove it without going through the whole process they have now, that takes forever, we would be much better off. (P12)

When considering solutions to challenges of timeliness, participants stressed that the willingness of elected officials to implement policy to improve systems is the most important factor for improvement. Mortality information systems are a low priority for improvement compared to other information systems that are more lucrative to the state. Participants shared that the expectation is that data systems are "good enough" to function, and so it is taken for granted that they will continue to function without any meaningful improvement. The lack of recent updates to policy surrounding mortality information systems was cited by participants as a sign of the lack of interest for improvement from politicians. As one 2013 report by NAPHSIS poignantly summarizes, "Typically, unless there is a crisis of availability, federal policymakers are not inclined to intervene.[73]"

6.1.3 *Archaic Government Data Systems.*

The utility of mortality information systems is, in part, limited by government organizations' poorly optimized and slowly adopted technical systems. We had expected archaic data systems to be a primary contributing factor to mortality information system utility. Participants certainly cited poor technical systems as a challenge that needs intervention. However, participants stressed that these technical issues were symptomatic of deeper policy challenges.

Interviewees said that the electronic systems used to manage, archive, and transfer death data are procedurally inefficient, technically slow, and difficult for researchers to navigate – primarily because of the old age of those systems. The specific systems participants identified as inefficient, slow, and difficult to navigate, varied depending on the participant group. For example, clinical researchers focused on the difficulty of navigating the user interface of the NDI system they were applying to access data from. Meanwhile, funerary and medical professionals noted the onerous user interface of the EHR and hospital charting software. State and federal government employees focused on the EDRS and the difficulty employees encounter when using search features to help them match the identity of decedents.

All participants noted the old age of the information systems they interfaced with, identifying the archaic technical architecture as a barrier to promptly processing or accessing mortality data. One participant noted that the last time the software they were using had been updated was 15 years previously – when the digital system was updated to be a basic archival system supplementing a paper system (with the paper system still being primary even after the update). Another participant similarly noted that, in their experience, mortality information systems persisted as paper systems long after federal vital statistics systems were updated to more robust digital systems.

Additionally, all participants favored updating and advancing electronic mortality information systems. Participants noted that COVID-19 had provided an opportunity for systems to be updated – although updates were piecemeal and only to support specific problems related to accounting for the sheer number of dead. All participants reported that the updates made during COVID-19 were insufficient. Multiple participants pointed out a clear disparity, asking – if everyone agrees that information systems should be updated, and other vital records systems have been able to be updated, why have mortality data systems lagged behind?

One participant attributed the disparity to the difficulty of disrupting the existing budgetary and organizational status quo regarding the mortality data ecosystem in particular:

> A lot of people believe in the advancement of these data systems. The technology is there. But there is a fear that if anyone tries, it will upend the whole infrastructure that has been created over the last 40 years. The coordination of federal and state levels depends on this fragile system that is locked in place at this point and no one wants to stick their neck out to be the one to take on all of the logistical issues in modernizing. (P06)

In our conversations with participants about archaic technical systems, we saw echoes of previous HCI research that has found that government organizations rely heavily on legacy systems to support their business-critical functions [2]. However, whereas previous literature has identified legacy systems as technical barriers to government functions, our participants stressed that the red tape of rules and procedures was the main barrier to updating technical systems. Participants stressed that even when superior technical systems were available, political factors prevented organizations from prioritizing updates.

6.1.4 Lack of Data Quality (Lack of Consistent, Complete, and Accurate Data). When reflecting on the challenges of utility (e.g., the difficulty researchers have in using mortality data, even if they can access an NDI file), participants spoke about the quality challenges of the data they have access to. Interviewees shared that consistent, complete, and accurate data are imperative for researchers to have access to, to serve the public good. However, data quality problems persist, predominantly due to un-standardized data management rules and procedures.

Lack of consistent and complete data. Participants shared that the consistency and completeness of data mostly run into challenges during the data collection phase at the clinical level (e.g., the moment of death certificate creation):

> It's not something that a lot of people want to hear, but the death certificate is often wrong. Sometimes people die under mysterious circumstances, sometimes there is no loved one to confirm identity, sometimes things just happen in the moment of death that cause information to be recorded wrong. The death certificate itself is not a guarantee of accuracy, it's merely the best they can discover at the time. (P08)

Another participant shared:

> We are not able to improve the accuracy of data until we have a standardization of methods and tools. Honestly, the content fields of the death certificate is pretty

> universal across states and hospitals. The problem is that the way it is filled out, who fills it out, and what fields should be filled out varies hospital by hospital and patient by patient, depending on what is happening on the floor... There is a real need for federal standardization of methods for how these things get filled out, and training for physicians. (P07)

Many participants used the data quality of mortality data during COVID-19 as exemplary of broader data quality issues:

> In COVID, we did not have accurate death certificate records... we were strictly going off of cremation numbers, and that led to a lot of issues for researchers trying to figure out accurate epidemic trends. (P16)

Interviewees shared that the completeness of data also runs into challenges when trying to access data at the NDI level. The NDI does not give researchers access to all death certificate data (e.g., researchers do not have access to contributing factors of someone's death, only the cited cause of death). Meaning that even when more complete data is available in NDI records, researchers cannot access the entirety of the record:

> One of the biggest issues researchers like us run into is that we don't have access to data on contributing things to someone's death, only the cited cause of death. (P03)

Another participant added:

> When a death happens, we most often just get metadata of the death certificate, like the header record. We would love to get additional data all the time – like who the reporter was, when each of the death certificate fields were filled, and by who. (P06)

The main reason the NDI cannot give researchers access to certain pieces of data is that each state's legislation independently determines which information in their state's records can be shared with researchers. For example, Tennessee may limit the information accessible to researchers to the primary cause of death, and not allow access to contributing causes. On the other hand, Washington may allow for both the primary cause of death and all contributing causes of death to be reported.

Each state sets its own rules for access and systems to apply for access – creating logistical barriers for researchers seeking complete data. Interviewees said that variable legislation means critical information about the mortality data is only sometimes evident. Often, the answers to the following questions vary across states and systems:

(1) Is the mortality data available?
(2) What mortality data is available?
(3) Who is the mortality data available to?
(4) Under what conditions is certain mortality data available?

Regarding variable legislation, one participant shared:

> The fact that these data are so siloed state by state aren't really helping at all... in the late 70s there wasn't any legislation. Then the CDC says data is relevant so they went around to various data jurisdictions and then structured this loose confederation of vital statistics offices, voluntary at the state level. Then the files were accumulated via paper at the state level without any federal policy. To keep things simple the CDC essentially created a self-funding program at the state level that allows the NDI to exist without much outside legislative mandate. Each state does their own thing however they want to do it... That variability makes the whole system impossible to make coherent or standardize. (P08)

Lack of accurate data. Beyond incomplete death certificate data and an inability to access data even when it is complete, data inaccuracy is an issue for researchers. The data in the NDI is susceptible to being wrongly matched or otherwise wrongly linked to decedents during the matching/identity linkage cleaning processes. In other words, researchers cannot be sure they have the most accurate data because government offices are not sure they have matched it accurately, and there is no way for a researcher to double-check. As one participant summarized:

> Federal agencies, god bless them, try their best, but data linkage is a constant issue for researchers. Especially when doing clinical trials, we need to get the right person, be able to rule out people with the same name, and ensure dead or not dead validity. Matching is difficult. Many people have the same names, many have immigrated... if we don't have accurately matched data it can cause harm and make our job much more difficult later on. (P06)

Participants shared that matches in the NDI can have a high probability of being wrong:

> Names are common across many people... its hard to triangulate the date of birth with the social security number with the age, race, gender, state of birth, etc... [the people matching] count on either all the variables matching or a statistics probability algorithm that says that this is probably a match. But information changes over time. People change their name over time. People change their age all the time. The match is not by any means a guarantee. (P04)

Participants were quick to stress that the high probability of wrong matches was not due to the skill or capacity of the employees of the CDC and NDI offices – in fact, participants had universal praise for the expertise of the employees at the NDI. Instead, participants said that law and policy were the main barriers to quality data matching. Exemplifying this: the main solution participants suggested to fix data quality issues was updating the Model State Vital Statistics Act, which most participants called a "Model Law." Participants noted that the 1992 Model State Vital Statistics Act is outdated and did nothing to support effective data quality practices. For participants, a more effective Model Law, implemented in policy by the NCHS, would standardize procedures and definitions across organizations and states, considering modern technology considerations. Participants also stressed that a Model Law should support uniformity by enshrining financial incentives for data quality practices in federal law (P02, P04).

When asked whether there were examples of successful standardization in the ecosystem, multiple experts referenced a policy revision passed in 2003 by the federal government. In 2003, the federal government updated the federal 1989 birth and death certificate standards to implement a Standard Certificate, which includes much more information about a patient's cause of death than previous standards [73]. The 2003 revision attempted to standardize the data that was collected at the hospital level and allowed for more information about the cause of death of patients to be available in databases such as the NDI. Experts said that this policy was successful because improving the certificate was specific, actionable, and was to the benefit of multiple organizations (e.g., the standard certificate helped hospitals streamline their recordkeeping).

6.2 Cultural Aversion to Engaging with Death

In the previous section, we demonstrated that participants believed more consistent and coherent policies could improve efficiency and access issues with mortality data. However, participants also shared discouragement about developing and implementing policy changes due to a cultural aversion to engaging with or discussing death – leading to policy-makers being unwilling and disincentivized to drive change. In this section, we describe what our participants considered to be the impact of cultural aversion to death on implementing policy solutions to bottlenecks and inefficiencies in the mortality data ecosystem.

In addition to policy, participants said that challenges with budget, lagging technical upgrades, and poor staffing that drove inefficiencies, were symptoms of a system of people with power who do not want to confront death (P03, P06, P08, P12). P08 summarized this phenomenon,

> No one wants to touch death data – it's spooky. It's not happy like a certificate for a baby's birth or a marriage certificate. It's dark, doesn't make money, and depressing. For policymakers, it won't get you reelected like tax breaks will. Who in their right mind would want to touch death if they don't have to?

Participants characterized cultural aversion to engaging with death as leading to an aversion in organizations to put energy, money, or overhead into optimizing policy to improve the data ecosystem. P10 highlighted,

> The U.S. is bizarre. We do not talk about death until we have to. Maybe people are scared... But either way, that aversion in the culture trickles down into the aversion of these organizations to coordinate, or to even be given the resources from the powers that be to optimize that coordination.

Multiple participants mentioned the trickle-down effect of cultural aversion to discussing death as a main barrier to government organizations' funding. Per P11, "Without normalizing discussing death in our own homes, how will we normalize the government discussing it? And if the government won't discuss it, how will they fund improvements to it?"

Other participants connected the cultural aversion to discussing death with policymakers' aversion to promoting policy change to their constituents. P12 described the phenomenon, saying "No one at the local level wants to hear a congressman campaign on promises to optimize mortality data infrastructure. It just isn't attractive... Voters hate hearing about death." P04 added, "It's career suicide to talk about death to voters."

Participants discussed specific areas where concrete actions could not be taken due to cultural aversion. For example, one participant attributed the lag in Colorado policymakers passing Provision 25-2-117(1)[25] – which clarified state processes for the disclosure of vital records – to a fear of politicians being stigmatized for championing the revision of the provision. Another participant attributed the relative lack of funding provided annually to the NDI by the state of Alaska to politicians not seeing a financial benefit to discussing death with their constituents.

Our finding that cultural aversion impedes coordination in the mortality data ecosystem needs more data to be generalized beyond our participant's experiences and perspectives. However, our participants stressing that coordination around mortality data does not exist in a cultural vacuum is notable when considering the day-to-day processes of how policy is made and revised by policymakers. We further discuss the implications of cultural aversion to death to mortality data coordination in Section 7.2.

7 Discussion

The challenge of how to improve the mortality data ecosystem is fundamentally a CSCW challenge: how do we design a large-scale information system with diverse kinds of data work and variable standards in need of more coherent policies to guide coordination? The mortality data ecosystem is a case of a massive information system consisting of policy and cultural 'baggage' constructed over many years and political administrations.

While our study maps out the ecosystem and qualitatively identifies challenges, articulating specific fixes (e.g., legal recommendations) is beyond the scope of the current work and will require more focused work. Instead, we provide guiding principles that will support that future work. Our discussion first describes these principles, and then considers how cultural relationships to death in the U.S. impact policy implementation. We conclude by discussing the impacts of data-sharing policies (e.g., HIPAA) on researcher access to mortality data. For each section, we suggest paths forward for future work.

7.1 Guiding Principles

Based on our study findings, we suggest several guiding principles for future work in the mortality data ecosystem.

Principle 1: Take an ecosystems view to mortality data. We increasingly depend on data moving between systems that were developed without any knowledge or consideration of one another. In the case of the mortality data ecosystem, we see a massive web of technical, organizational, and policy systems developed independently across decades. As we found in the bottlenecks around data linking procedures, state and federal data linking are typically redundant. The redundancy is because data linkage law at the state level has been developed and enshrined independently of federal law. In other words, a state like California has developed its data linkage process over the last 100 years, simultaneously with the federal government but not coordinated with the federal government.

If designers only focus on one piece of this ecosystem (e.g., focus on either state policy or federal policy), they will miss an important piece of the puzzle. Missing a portion of the whole picture will limit the effectiveness of any interventions designers propose. As we have shown in this study, the whole is more than a sum of its discrete parts – the larger ecosystem is a series of interdependencies between those parts. An ecosystems approach is critical to addressing the underlying data coordination issues. An ecosystems approach can help identify opportunities for improvement while also attending to how responsibilities and resources are distributed across the ecosystem.

More recently, CSCW has suggested that an ecosystems approach is a useful approach for finding solutions to coordination issues in large-scale social, political, and algorithmic systems [105, 62, 10]. Perhaps most relevant for the policy-based ecosystem challenges in the mortality data system, one CSCW model with special utility is the policy knot model by Jackson et al. [56]. When considering the complex interdependencies of technical, social, and policy forces, Jackson et al.'s policy knot is a useful analytic lens due to its explanatory power about the inseparability of design, practice, and policy. As we previously introduced in 3.3, policy knots are "complex interdependencies... that collectively determine the shape, meaning, and trajectory of shifting computational forms [56]." Our findings suggest that the mortality data ecosystem is a case of a policy knot. The policy knot model responds to the tendency of HCI researchers and practitioners to silo the forces that impact technology usability – and overlook how different forces (policy and design forces, for example) are inseparable from one another. In the mortality data ecosystem, we see policy and design as interdependencies that should be addressed together.

A policy knot model can be especially helpful in addressing the challenges of mortality data records evolving over time. The constraints of policy (hospitals, state reporting, and federal reporting having different regulatory standards for the ongoing updating of records) do not match the design affordances of the database technologies or the practice needs of data workers. Using a policy knot model to compare the interdependencies of policy, design, and practice can help designers identify the most pressing challenges for coordination across the ecosystem. In this case, the policy knot model helps to identify that variable regulations need to be brought in line with the concrete realities of data workers. After variable regulations are brought into line, technical design for databases that account for evolving data records can be implemented more effectively. For example, we imagine that by bringing data processing policy in line with the day-to-day practices of spreadsheet cleaning, features can be improved to standardize workflows and reduce lag.

Principle 2: Support under-resourced phases in the mortality data flow through the development of standards. While federal-level standards support some parts of the mortality data flow (e.g., the data on a death certificate), other parts have been historically under-resourced. Identifying these under-resourced areas and developing standards to support them will reduce costs and errors in the long term. For example, one major cause of lag is a lack of reporting standards across state jurisdictions. State reporting is under-standardized – with little federal oversight over when and how the data gets reported. Although states are expected to report their data annually, the policy beyond that expectation is variable and ineffective. Greater standardization in that particular phase of the data flow (EDR to NVSS) could go a long way to reducing data lag throughout the entire system. CSCW should prioritize supporting under-standardized phases in the data flow.

Standardizing data flows to help identify and design solutions for under-standardized areas is a focus of CSCW scholarship generally [7, 69], and CSCW healthcare scholarship specifically [40, 11]. Of particular note is the seminal work by Winthereik and Vikkelsø on the role that ICTs can play in standardizing data artifacts [104]. Winthereik and Vikkelsø argue that ICTs can improve data quality and timely data processing through articulation work, demarcating areas of responsibility, and identifying specific places within and between organizations in which coordination and integration should happen [104]. Designers can then prioritize those areas for intervention.

Principle 3: Focus on outcomes over processes when supporting distributed work. When possible, interventions should focus on outcomes rather than specific processes. During our analysis, we found places where the work is unified (e.g., NDI identity matching) – where the standardization of processes is well-poised to improve quality and speed. However, federal-level standardization of processes is not always preferable. For example, in the most distributed parts of the data flow (e.g., local reporting and state-level vital records), data collection processes vary but can still be timely. These processes' differences should be respected, so long as they achieve common outcomes. To ensure interoperability, standards in these cases should focus on the outcomes – however they are achieved – rather than dictating processes.

Unifying outcomes might best be supported by some unification of documentation practices. Sambasivan et al. previously found that improving cross-organizational documentation can improve coordination in distributed workflows [85]. This finding certainly applies here. For example, one can imagine multiple organizations in the mortality data ecosystem coordinating around the common goal of improving cross-organizational documentation, without necessarily needing to standardize their processes.

Principle 4: Consider the discrepancies between responsibilities and beneficiaries. Our interviews revealed no common incentives for policy or process to change. For example, we found that when it came to organizations hiring additional employees to do the data work or updating their archaic technical systems, they did not do so because there was no meaningful incentive – financial or otherwise. The resistance makes sense. After all, states would not want to purchase new software the federal government has requested if the cost comes at their expense and does not provide them any clear benefit. Designers contemplating policy interventions should prioritize asking: who do policy solutions benefit? Who might those policy solutions harm? And how can policy account for discrepancies between responsibilities and beneficiaries of standardized policies? Considering these questions, designers should focus on shifting labor around in ways that better align with benefits.

When considering the discrepancies between responsibilities and beneficiaries, our findings suggest that examining data work around mortality data in healthcare – and the infrastructures that support it – is an important area of future research. While data work was not the focus of our inquiry, when considering the bottlenecks in the mortality data ecosystem it is clear that significant labor is required to clean and process data at the medical, local government, and federal government levels. After all, policy decisions do not just impact end-users like researchers – they also impact the individuals doing the labor throughout the ecosystem [30]. A large-scale analysis of data work in the mortality data ecosystem would be helpful follow-up research to more comprehensibly understand the intersections of policy and data work in the healthcare sector. To frame this future data work analysis, our findings about the importance of institutional structures in shaping data flow lead us to believe that researchers should additionally consult infrastructure studies focused on the ecologies of data work in healthcare [12, 80].

7.2 A Case of Charismatic Data: How Do We Account for Cultural Aversion to Death?

Through our analysis, we found that a cultural aversion to engaging with death leads to an avoidance of policymakers to develop policy solutions for bottlenecks and inefficiencies in the mortality data ecosystem. Participants discussed how cultural aversion resulted in concrete barriers to politicians championing policy (e.g., Colorado Provision 25-2-117(1)[25]) and disincentivizing budget allocations (e.g., Alaska maintaining a low level of financial support for the NDI). We were surprised by participants citing cultural challenges in these medical and governmental contexts as underlying issues. We had expected technical and organizational barriers, but had not considered the momentous constraints that cultural relationships to death and its data might put on the ecosystem. Although death denial in the U.S. is a well-documented phenomenon [99, 59], the challenge that cultural relationships with death impose on governmental information systems is a new challenge for CSCW to consider.

In addition to prioritizing the principles above, we argue that designers should prioritize addressing how a cultural aversion to engaging with death is a barrier to coordination in the mortality data ecosystem. We separate this challenge from the policy challenges and principles because of a clear lack of precedent in related literature connecting cultural narratives around death to information ecosystems. The design space around a cultural aversion to death in the mortality data ecosystem is non-existent, and baby steps should be taken by CSCW research to scope the problem and develop solutions. In the spirit of providing one direction designers can begin to follow, we argue that using a charismatic data framework [79] can help designers establish a well-scoped design space.

Pine and Liboiron have argued that data can be charismatic, inspiring devotion so strong that it moves an audience to action [79]. They use this term to describe data with values, judgments, and fears that are already culturally salient and have the potential to launch action in ways other data may not. Pine and Liboiron leverage this idea of charismatic data to discuss ways that measurements get their power to move people to action. They cite the examples of blood (which signals fear and crisis), feces (which historically signals disgust and danger, and child abuse (which signals violence), as measurements that have a terrifying charisma. Pine and Liboiron argue that the terrifying charisma of blood, feces, and child abuse have the potential to launch action in ways that other data may not. Additional work has used the concept of charismatic data to discuss the ways that the potential charisma of food data and mental health data can influence how organizations choose to share that data with their stakeholders [26, 101].

We believe that mortality data is also a kind of charismatic data – a data that signals fear, crises, disgust, danger, and violence – that has the potential to launch individual and organizational action. We suggest that designers should employ the charismatic data framework when designing for more standardized data measurement in the mortality data ecosystem.

Although previous end-of-life HCI research has not formally labeled mortality data as "charismatic," research has highlighted several ways that the values, judgments, and fears already salient in a culture around death inform decisions around personal data management at the end of life [46, 76]. We have found that – similar to work on personal data – mortality data policy is also influenced by cultural values, judgments, and fears around death. Like the charismatic data examples identified by [79] (i.e., blood, feces, and child abuse), it follows that mortality data would also come with a terrifying charisma that compels systems to launch action.

But there is a problem: In the mortality data ecosystem, the terrifying charisma of mortality data impedes action. Contrary to Pine and Liboiron's examples, here, the charisma of the data disincentivizes action/coordination (e.g., the politicians who will not champion system reform because of public blowback to discussing the topic of death). The design question becomes: How do designers reposition a type of charismatic data that impedes coordination to a type of charismatic data that facilitates improved coordination? The formulation of this question is helpful for designers in of itself – helping to scope the social-technical challenge.

Pine and Liboiron's charismatic data framework presents one possible solution to challenges – the power of measurement. The authors argue that the act of naming data as charismatic data allows for a more intentional measurement of data by mapping the data into preexisting values. As a "technique of definition that distills something amorphous into something essential [79]," more intentional measurement of data can identify and address root causes of challenges. Naming mortality data as a discrete type of data – charismatic data – can help designers put boundaries around it, allowing for a more targeted measurement and intervention.

One example of how this naming can be helpful for designers is Pine and Liboiron's argument that boundarying can shape prioritization at the administrative level. For example, previous research has shown that measuring charismatic data can lead to more effective evaluations of medical practice, organizational performance, and inter-unit coordination [79, 101, 26]. We imagine that by viewing mortality data as charismatic data, designers can improve organizational performance and inter-unit coordination through more targeted administration prioritization. We imagine that identifying mortality data as having a terrifying charisma might also drive more open and honest conversations about death at the organizational level.

How to most effectively design large-scale healthcare data ecosystems that account for cultural aversion to certain types of data is a complex and open question. The question of who should be responsible for designing systems that address such a widespread social challenge, like cultural aversion, is also a question without a clear answer. We suggest CSCW researchers and designers can play a critical role in addressing those challenges – or at least provoking new and nuanced conversations about those challenges. We offer charismatic data and the power of measurement as one framework already present in CSCW literature that might be applied to begin to design a more effective mortality data ecosystem. Additionally, this framework is one example of insight already present in CSCW that can be an ignition point for urgent interdisciplinary conversations about the intersections of culture, policy, and healthcare data.

7.3 Impacts of Data-Sharing Policies on Access

Another aspect that designers must consider is policy's role in shaping activities and the various principles we have outlined above. Our participants extolled the impacts of costs on the access to mortality data, especially data that has not been de-identified. The Health Insurance Portability and Accountability Act (HIPAA) is the U.S. policy that maintains the national standards to protect individuals' medical records and other identifiable health information [96]. An aspect of these policies outlines how identifiable health data should be handled and stored. The only way researchers are considered a "covered entity" under this policy is if they are also health care providers who electronically transmit health data [98]. Thus, most researchers have to affiliate as a Business Associate with the covered entity to (1) go through a distinct process to identify "need to know" with the covered entity regarding the data and (2) formalize the business associate being directly liable for penalties if there is a HIPAA violation [98].

HIPAA violations can range anywhere from $100 to $50,000 per violation depending on the type of civil violation that took place [50]. The cost to manage protected health data, let alone share it externally, is a hidden cost that impacts access to the data. Thus, many organizations are continually conducting cost/benefit analyses with respect to sharing data. With respect to mortality data, this is still viewed as personal health data in the U.S. for 50 years following the date of the individual's death [96]. Thus, in the eyes of the U.S. government, there is really no difference between your health data and whether you are alive or have passed away – a critical aspect that designers must consider. Local policies also shape the ability to efficiently share health data. Stakeholders in under-resourced contexts are often hesitant or slow to share health data [3], which makes sense considering the hidden costs associated with data and data sharing.

One way to ensure mortality data is available is the increasingly popular notion of data donation after death. This addresses issues related to both HIPAA and GDPR in the European Union [89]. While this is not currently a commonplace aspect of mortality or preparations leading up to a death event, it is a provocative point of consideration that addresses issues of who owns the data to share and making it more accessible in general. However, this further complicates the mortality data ecosystem as it creates a new – less formally regulated – access point to data.

8 Limitations

This study has several limitations. First, though when designing this study we considered mapping the data ecosystem of both de-identified and identified data, we quickly found that these types of data followed quite different data journeys, flowing through unique ecosystems, organizations, and databases. Having to choose which ecosystem to start with so we could retain a manageable research scope, we decided this first study should focus on identified data due to its utility to clinical researchers. We strongly believe that a follow-up study should be conducted mapping the de-identified mortality data ecosystem and comparing the findings to this study's findings.

This study is additionally limited by its strict focus on researcher access to mortality data. Our findings are not necessarily generalizable to other use cases of U.S. vital records information systems (such as researcher access to birth data). By focusing on researcher access to mortality data, we scoped our study on a smaller piece of a vast U.S. vital records data ecosystem. Additionally, mortality data have more end-users than the clinical research end-users we have focused on. For example, the private sector (e.g., pharmaceutical companies) is a quickly growing end-user of mortality data. Although we imagine that similar access challenges exist for other end-user groups, our study is not necessarily generalizable to the needs and challenges of those other non-researcher end-users.

Finally, this study focuses on instances where an individual dies within a hospital system and is reported dead within a health facility. However, deaths occur outside of health systems as well [63, 58, 106]. For example, if someone dies in at-home hospice care, or in a sudden car crash, they do not need to be transported to the ER to be declared dead. The literature we reviewed, and interviews we conducted, did not highlight morgue records and police records as part of the rest of the mortality data ecosystem, even though those records are places where mortality data is archived and processed. Mortality data generated outside of a health facility is a blind spot for this study and, like de-identified data, we strongly believe a follow-up study is needed to make sense of this related data ecosystem.

9 Conclusion

Each of us will one day have data about our deaths encoded into the mortality data ecosystem. Ultimately, this study asks: How do we want that data to be managed?

The research presented here focuses on improving the mortality data ecosystem for researchers accessing critical information that can be turned into life-saving medical interventions. We argue that to more effectively design coordinated systems that support access, technical solutions are not enough – we must improve policy. To that end, we mapped the mortality data ecosystem and conducted an analysis identifying emergent symptoms of access barriers (cost, timeliness, archaic technical systems, data quality) and their underlying causes (policy and cultural aversion to engaging with death). Based on our analysis, we have argued for principles to be used to guide future work.

Our hope is that future research can operationalize these principles in design solutions. Further, we hope that this study highlights the necessity for CSCW researchers and practitioners to incorporate policy affordances and constraints in their design considerations – especially when designing for big-data healthcare ecosystems. Our findings lead us to believe that much more research is needed on the cultural relationship of death as it relates to information system design and healthcare data management. Finally, we argue that future work on U.S. vital records information systems should take special interest in the impacts of data-sharing policies on researcher data access.

Acknowledgments

We first thank our anonymous reviewers, whose close reading and feedback strengthened this manuscript considerably. We additionally thank our participants, who shared their stories and experiences with us over the course of this study. We thank the colleagues who helped us refine our thought processes and approaches through discussion. In particular, we thank Amy O'Hara, Isabella Chu, and the members of the Identity Lab for their guidance. Finally, we thank the National Science Foundation for their ongoing support of this work.

References

[1] Ryan M Ahern et al. "Improving the public health utility of global cardiovascular mortality data: the rise of ischemic heart disease". In: *Population health metrics* 9 (2011), pp. 1–11.

[2] Assia Alexandrova, Lucia Rapanotti, and Ivan Horrocks. "The legacy problem in government agencies: an exploratory study". In: *proceedings of the 16th annual international conference on digital government research*. 2015, pp. 150–159.

[3] E. Anane-Sarpong et al. "You cannot collect data using your own resources and put It on open access": Perspectives from Africa about public health data-sharing". In: *Developing world bioethics* 18.4 (2018), pp. 394–405. DOI: 10.1111/dewb.12159.

[4] Marıa P Angel and Danah Boyd. "Techno-legal Solutionism: Regulating Children's Online Safety in the United States". In: *Proceedings of the Symposium on Computer Science and Law*. 2024, pp. 86–97.

[5] Carl V Asche et al. "Evaluation of healthcare interventions and big data: review of associated data issues". In: *Pharmacoeconomics* 35 (2017), pp. 759–765.

[6] Carol Bacchi. *Analysing policy*. Pearson Higher Education AU, 2009.

[7] Carlo Batini, Monica Scannapieco, et al. "Data and information quality". In: *Cham, Switzerland: Springer International Publishing* (2016).

[8] Lehn M Benjamin, Amy Voida, and Chris Bopp. "Policy fields, data systems, and the performance of nonprofit human service organizations". In: *Human Service Organizations and the Question of Impact*. Routledge, 2020, pp. 73–92.

[9] Jules J Berman. "Confidentiality issues for medical data miners". In: *Artificial intelligence in medicine* 26.1-2 (2002), pp. 25–36.

[10] Sander Bogers et al. "A situated exploration of designing for personal health ecosystems through data-enabled design". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. 2018, pp. 109–120.

[11] Claus Bossen and Randi Markussen. "Infrastructuring and ordering devices in health care: Medication plans and practices on a hospital ward". In: *Computer Supported Cooperative Work (CSCW)* 19 (2010), pp. 615–637.

[12] Claus Bossen et al. “Data-work in healthcare: The new work ecologies of healthcare infrastructures”. In: *Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion*. 2016, pp. 509–514.

[13] Taxiarchis Botsis et al. “Secondary use of EHR: data quality issues and informatics opportunities”. In: *Summit on translational bioinformatics* 2010 (2010), p. 1.

[14] Virginia Braun and Victoria Clarke. “Using thematic analysis in psychology”. In: *Qualitative research in psychology* 3.2 (2006), p. 77.

[15] Erin G Brooks and Kurt D Reed. “Principles and pitfalls: a guide to death certification”. In: *Clinical medicine & research* 13.2 (2015), pp. 74–82.

[16] Jed R Brubaker and Gillian R Hayes. “" We will never forget you [online]" an empirical investigation of post-mortem myspace comments”. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 2011, pp. 123–132.

[17] Jed R Brubaker, Gillian R Hayes, and Paul Dourish. “Beyond the grave: Facebook as a site for the expansion of death and mourning”. In: *The Information Society* 29.3 (2013), pp. 152–163.

[18] Jed R. Brubaker and Vanessa Callison-Burch. “Legacy Contact: Designing and Implementing Post-mortem Stewardship at Facebook”. en. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. San Jose California USA: ACM, May 2016, pp. 2908–2919. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858254. URL: https://dl.acm.org/doi/10.1145/2858036.2858254 (visited on 05/05/2022).

[19] Jed R. Brubaker et al. “Stewarding a legacy: responsibilities and relationships in the management of post-mortem data”. en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto Ontario Canada: ACM, Apr. 2014, pp. 4157–4166. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557059. URL: https://dl.acm.org/doi/10.1145/2556288.2557059 (visited on 05/05/2022).

[20] Miriam C Buiten. “Towards intelligent regulation of artificial intelligence”. In: *European Journal of Risk Regulation* 10.1 (2019), pp. 41–59.

[21] Peter Byass. “Who needs cause-of-death data?” In: *PLoS Medicine* 4.11 (2007), e333.

[22] *CDC: Breaking Barriers for Drug Death Data*. URL: https://www.cdc.gov/surveillance/blogs-stories/Breaking-Barriers-for-Drug-Death-Data.html.

[23] *CDC: Vital Records Information Systems*. URL: https://www.nlm.nih.gov/nichsr/stats_tutorial/section3/mod4_records.html.

[24] Victoria Chui et al. “Applying Human-Centered Data Science to Healthcare: Hyperlocal Modeling of COVID-19 Hospitalizations”. In: *Companion Proceedings of the 2023 ACM International Conference on Supporting Group Work*. 2023, pp. 24–26.

[25] *Colorado Provision 25-2-117(1)*. URL: https://casetext.com/statute/colorado-revised-statutes/title-25-public-health-and-environment/vital-statistics/article-2-vital-statistics/section-25-2-117-certified-copies-furnished-fee.

[26] Shiva Darian et al. “Enacting data feminism in advocacy data work”. In: *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW1 (2023), pp. 1–28.

[27] Nancy Darling. “Ecological systems theory: The person in the center of the circles”. In: *Research in human development* 4.3-4 (2009), pp. 203–217. DOI: 10.1080/15427600701663023.

[28] Charles E Dean and Paul D Thuras. “Mortality and tardive dyskinesia: long-term study using the US National Death Index”. In: *The British Journal of Psychiatry* 194.4 (2009), pp. 360–364.

[29] Barry Dewitt and George W Torrance. “Incorporating mortality in health utility measures”. In: *Medical Decision Making* 40.7 (2020), pp. 862–872.

[30] Lynn Dombrowski et al. “E-government intermediaries and the challenges of access and trust”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 21.2 (2014), pp. 1–22.

[31] Dylan Thomas Doyle, Charlie Blue R Brahm, and Jed R Brubaker. “" I hate you. I love you. I’m sorry. I miss you." Understanding Online Grief Expression Through Suicide Bereavement Letter-Writing Practices”. In: *Proceedings of the ACM on Human-Computer Interaction* 8.CSCW1 (2024), pp. 1–27.

[32] Dylan Thomas Doyle and Jed R Brubaker. “" I Am So Overwhelmed I Don’t Know Where to Begin!" Towards Developing Relationship-Based and Values-Based End-of-Life Data Planning Approaches”. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–14.

[33] Dylan Thomas Doyle and Jed R Brubaker. “Digital legacy: a systematic literature review”. In: *Proceedings of the ACM on Human-Computer Interaction* 7.CSCW2 (2023), pp. 1–26.

[34] Dylan Thomas Doyle et al. “Stories That Heal: Characterizing and Supporting Narrative for Suicide Bereavement”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 18. 2024, pp. 354–366.

[35] Eric L Eisenstein et al. “Death: the simple clinical trial endpoint”. In: *Studies in health technology and informatics* 257 (2019), p. 86.

[36] RI Eldridge and PD Sutton. “Births, marriages, divorces, and deaths: Provisional data for December 2006”. In: *National vital statistics reports* 55.20 (2007).

[37] Gunnar Ellingsen, Morten Hertzum, and Line Melby. “The tension between national and local concerns in preparing for large-scale generic systems in healthcare”. In: *Computer Supported Cooperative Work (CSCW)* 31.3 (2022), pp. 411–441.

[38] Sebastian S Feger et al. “’Yes, I comply!’ Motivations and Practices around Research Data Management and Reuse across Scientific Fields”. In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW2 (2020), pp. 1–26.

[39] Casey Fiesler. “Lawful users: Copyright circumvention and legal constraints on technology use”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–11.

[40] Geraldine Fitzpatrick and Gunnar Ellingsen. “A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas”. In: *Computer Supported Cooperative Work (CSCW)* 22 (2013), pp. 609–665.

[41] Kyle J Foreman, Mohsen Naghavi, and Majid Ezzati. “Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death”. In: *Population health metrics* 14 (2016), pp. 1–9.

[42] Kyle J. Foreman, Moshen Naghavi, and Majid Ezzati. “Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death”. In: *Population health metrics* 14 (2016), pp. 1–9. DOI: 10.1186/s12963-016-0082-4.

[43] Katie Z. Gach and Jed R. Brubaker. “Experiences of Trust in Postmortem Profile Management”. In: *ACM Transactions on Social Computing* 3.1 (Feb. 2020), 2:1–2:26. ISSN: 2469-7818. DOI: 10.1145/3365525. URL: http://doi.org/10.1145/3365525 (visited on 05/05/2022).

[44] Katie Z. Gach and Jed R. Brubaker. “Getting Your Facebook Affairs in Order: User Expectations in Post-mortem Profile Management”. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (Apr. 2021), 174:1–174:29. DOI: 10.1145/3449248. URL: http://doi.org/10.1145/3449248 (visited on 05/05/2022).

[45] Jonathan Grudin. “Why CSCW applications fail: problems in the design and evaluationof organizational interfaces”. In: *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*. 1988, pp. 85–93.

[46] Rebecca Gulotta et al. “Legacy in the age of the internet: reflections on how interactive systems shape how we are remembered”. In: *Proceedings of the 2014 conference on Designing interactive systems*. DIS ’14. New York, NY, USA: Association for Computing Machinery, June 2014, pp. 975–984. ISBN: 978-1-4503-2902-6. DOI: 10.1145/2598510.2598579. URL: http://doi.org/10.1145/2598510.2598579 (visited on 05/05/2022).

[47] Randy Hanzlick. “Quality assurance review of death certificates: a pilot study”. In: *The American journal of forensic medicine and pathology* 26.1 (2005), pp. 63–65.

[48] *Health IT Analytics: Understanding De-Identified Data, How to Use It in Healthcare*. URL: https://healthitanalytics.com/news/understanding-de-identified-data-how-to-use-it-in-healthcare.

[49] National Center for Health Statistics (US). “Health, United States, 2010 with special feature on death and dying”. In: (2011).

[50] *HIPAA Violactions Enforcement*. URL: https://www.ama-assn.org/practice-management/hipaa/hipaa-violations-enforcement.

[51] Michael Jeffrey Daniel Hoefer et al. “The multiplicative patient and the clinical workflow: Clinician perspectives on social interfaces for self-tracking and managing bipolar disorder”. In: *Designing Interactive Systems Conference 2021*. 2021, pp. 907–925.

[52] Ryan A Hoffman et al. “Improving validity of cause of death on death certificates”. In: *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2018, pp. 178–183.

[53] *Hopkins Medicine: Definition of De-Identified Data*. URL: https://www.hopkinsmedicine.org/institutional-review-board/hipaa-research/de-identified-data.

[54] Chung-Yuan Hu et al. “Assessing the utility of cancer-registry–processed cause of death in calculating cancer-specific survival”. In: *Cancer* 119.10 (2013), pp. 1900–1907.

[55] Allaouzi Imane and Ben Ahmed Mohamed. “Multi-label categorization of french death certificates using nlp and machine learning”. In: *Proceedings of the 2nd international conference on big data, cloud and applications*. 2017, pp. 1–4.

[56] Steven J Jackson, Tarleton Gillespie, and Sandy Payette. “The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing”. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 2014, pp. 588–602.

[57] Sarah Charlotte Johnson et al. “Public health utility of cause of death data: applying empirical algorithms to improve data quality”. In: *BMC medical informatics and decision making* 21.1 (2021), pp. 1–20.

[58] William B Kannel and Arthur Schatzkin. “Sudden death: lessons from subsets in population studies”. In: *Journal of the American College of Cardiology* 5.6 (1985), 141B–149B.

[59] Allan Kellehear. “Are we a ‘death-denying’society? A sociological review”. In: *Social science & medicine* 18.9 (1984), pp. 713–721.

[60] Joshua A Kroll. “Outlining traceability: A principle for operationalizing accountability in computing systems”. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, pp. 758–771.

[61] Dhanunjaya R Lakkireddy et al. “Improving death certificate completion: a trial of two training interventions”. In: *Journal of General Internal Medicine* 22 (2007), pp. 544–548.

[62] Emily G Lattie et al. “Designing mental health technologies that support the social ecosystem of college students”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–15.

[63] Mary Elizabeth Lewis et al. “Estimated incidence and risk factors of sudden unexpected death”. In: *Open heart* 3.1 (2016).

[64] Yunji Liang, Xiaolong Zheng, and Daniel D Zeng. “A survey on big data-driven digital phenotyping of mental health”. In: *Information Fusion* 52 (2019), pp. 290–307.

[65] Marian F MacDorman and Sharon Kirmeyer. “Fetal and perinatal mortality: United States, 2005”. In: (2009).

[66] Michael Massimi et al. “Matters of life and death: locating the end of life in lifespan-oriented hci research”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’11. New York, NY, USA: Association for Computing Machinery, May 2011, pp. 987–996. ISBN: 978-1-4503-0228-9. DOI: 10.1145/1978942.1979090. URL: http://doi.org/10.1145/1978942.1979090 (visited on 05/05/2022).

[67] Colin D. Mathers and Dejan Loncar. “Updated projections of global mortality and burden of disease, 2002-2030: data sources, methods and results”. In: *Geneva: World Health Organization* 10 (2005).

[68] Wendy Moncur and David Kirk. “An emergent framework for digital memorials”. In: *Proceedings of the 2014 conference on Designing interactive systems*. DIS ’14. New York, NY, USA: Association for Computing Machinery, June 2014, pp. 965–974. ISBN: 978-1-4503-2902-6. DOI: 10.1145/2598510.2598516. URL: http://doi.org/10.1145/2598510.2598516 (visited on 05/05/2022).

[69] Eric Monteiro et al. “From artefacts to infrastructures”. In: *Computer supported cooperative work (CSCW)* 22 (2013), pp. 575–607.

[70] Elizabeth L Murnane et al. “Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management”. In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (2018), pp. 1–27.

[71] Kathryn A Myers and Donald RE Farquhar. “Improving the accuracy of death certification”. In: *Cmaj* 158.10 (1998), pp. 1317–1323.

[72] Mahin Naderifar, Hamideh Goli, and Fereshteh Ghaljaie. “Snowball sampling: A purposeful method of sampling in qualitative research”. In: *Strides in development of medical education* 14.3 (2017).

[73] *NAPHSIS Report: More Better Faster - Strategies for Improving the Timeliness of Vital Statistics - 2013*.

[74] *National Archives: Vital Records Information Systems Overview*. URL: https://www.archives.gov/research/vital-records#:~:text=Vital%20records%20most%20commonly%20refer,in%20the%20District%20of%20Columbia..

[75] *NCHS: Organization*. en. URL: https://www.cdc.gov/nchs/about/organization.htm.

[76] William Odom et al. “Technology heirlooms? considerations for passing down and inheriting digital materials”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’12. New York, NY, USA: Association for Computing Machinery, May 2012, pp. 337–346. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2207723. URL: http://doi.org/10.1145/2207676.2207723 (visited on 05/05/2022).

[77] Irene V Pasquetto et al. “Open data in scientific settings: From policy to practice”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 1585–1596.

[78] Kathleen H Pine and Claus Bossen. “Good organizational reasons for better medical records: The data work of clinical documentation integrity specialists”. In: *Big Data & Society* 7.2 (2020), p. 2053951720965616.

[79] Kathleen H Pine and Max Liboiron. “The politics of measurement and action”. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015, pp. 3147–3156.

[80] Kathleen H Pine et al. “Data work in healthcare: Challenges for patients, clinicians and administrators”. In: *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2018, pp. 433–439.

[81] Pipetius Quah, Andrew Li, and Jason Phua. “Mortality rates of patients with COVID-19 in the intensive care unit: a systematic review of the emerging literature”. In: *Critical care* 24 (2020), pp. 1–4.

[82] Matthew D Redelings, Matthew Wise, and Frank Sorvillo. “Using multiple cause-of-death data to investigate associations and causality between conditions listed on the death certificate”. In: *American journal of epidemiology* 166.1 (2007), pp. 104–108.

[83] Harry M Rosenberg. “Cause of death as a contemporary problem.” In: *Journal of the history of medicine and allied sciences* 54.2 (1999), pp. 133–153.

[84] Johnny Saldana. *Fundamentals of qualitative research*. Oxford university press, 2011.

[85] Nithya Sambasivan et al. ““Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI”. In: *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–15.

[86] Steven Schwartz. "The US Vital Statistics System: The Role of State and Local Health Departments". In: *Vital Statistics: Summary of a Workshop*. 2009.
[87] Christian P Selinger, Robert A Ellis, and Mary G Harrington. "A good death certificate: improved performance by simple educational measures". In: *Postgraduate medical journal* 83.978 (2007), pp. 285–286.
[88] Ping Shao et al. "Improving Real-World Mortality Data Quality in Oncology Research: Augmenting Electronic Medical Records With Obituary, Social Security Death Index, and Commercial Claims Data". In: *JCO Clinical Cancer Informatics* 7 (2023), e2300014.
[89] D.M. Shaw. "Defining Data Donation After Death: Metadata, Families, Directives, Guardians and the Route to Big Consent". In: *he Ethics of Medical Data Donation [Internet]* Chapter 10 (2019), pp. 1–9. DOI: 10.1007/978-3-030-04363-6_10.
[90] Dan Sholler. "Infrastructuring as an occasion for resistance: Organized resistance to policy-driven information infrastructure development in the US healthcare industry". In: *Computer Supported Cooperative Work (CSCW)* 29.4 (2020), pp. 451–496.
[91] Min-Woong Sohn et al. "Accuracy and completeness of mortality data in the Department of Veterans Affairs". In: *Population health metrics* 4.1 (2006), pp. 1–8.
[92] Anne Spaa et al. "Understanding the Boundaries between Policymaking and HCI". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–15.
[93] Susan Leigh Star and James R Griesemer. "Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39". In: *Social studies of science* 19.3 (1989), pp. 387–420.
[94] Jackie Street et al. "Sharing administrative health data with private industry: a report on two citizens' juries". In: *Health Expectations* 24.4 (2021), pp. 1337–1348.
[95] Moshen Naghavi adn Susanna Makela et al. "Algorithms for enhancing public health utility of national causes-of-death data". In: *Population health metrics* 8 (2010), pp. 1–14. DOI: 10.1186/1478-7954-8-9.
[96] *The HIPAA Privacy Rule*. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html#:~:text=The%20HIPAA%20Privacy%20Rule%20establishes,care%20providers%20that%20conduct%20certain.
[97] William W Thompson, Lorraine Comanor, and David K Shay. "Epidemiology of seasonal influenza: use of surveillance data and statistical models to estimate the burden of disease". In: *The Journal of infectious diseases* 194.Supplement_2 (2006), S82–S91.
[98] *To Whom Does teh Privacy Rule Apply and Whom Will it Affect?* URL: https://privacyruleandresearch.nih.gov/pr_06.asp#:~:text=Covered%20entities%20are%20defined%20in,which%20HHS%20has%20adopted%20standards.
[99] Laura Tradii and Martin Robert. "Do we deny death? II. Critiques of the death-denial thesis". In: *Mortality* 24.4 (2019), pp. 377–388.
[100] Stefanus Van Staden and Nicola J Bidwell. "Localised Trust in a Globalised Knot: Designing Information Privacy for Digital-ID". In: *ACM Journal on Computing and Sustainable Societies* 2.1 (2024), pp. 1–37.
[101] Amy Voida et al. "Competing currencies: Designing for politics in units of measurement". In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017, pp. 847–860.
[102] Nicole Gray Weiskopf and Chunhua Weng. "Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research". In: *Journal of the American Medical Informatics Association* 20.1 (2013), pp. 144–151.
[103] Cedric Deslandes Whitney et al. "HCI Tactics for Politics from Below: Meeting the Challenges of Smart Cities". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–15.
[104] Brit Ross Winthereik and Signe Vikkelsø. "ICT and integrated care: some dilemmas of standardising inter-organisational communication". In: *Computer Supported Cooperative Work (CSCW)* 14 (2005), pp. 43–67.
[105] Christine T Wolf et al. "The changing contours of" participation" in data-driven, algorithmic ecosystems: Challenges, tactics, and an agenda". In: *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2018, pp. 377–384.
[106] C Wren, J O'Sullivan, and C Wright. "Sudden death in children and adolescents". In: *Heart* 83.4 (2000), p. 410.
[107] Qian Yang et al. "The Future of HCI-Policy Collaboration". In: (2024).
[108] Angie Zhang et al. "Algorithmic management reimagined for workers and by workers: Centering worker well-being in gig work". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–20.

Received January 2024; revised April 2024; accepted May 2024