Qualitative data-mining (QDM), using the narrative data contained in child welfare case records, enables researchers to examine child welfare practice using relatively non-intrusive methods. QDM can increase our understanding of client populations and problems, child welfare worker actions, and case complexity. This paper reports on experiences from the Child Welfare Qualitative Data-Mining Project; outlines a seven-step guide to QDM methods; and describes how QDM can be used to enhance child welfare practice, research, and education.
In their daily practice, social service professionals routinely collect and record large quantities of data about client characteristics, practice interventions, and client outcomes (Epstein, 2002, 2009). While documentation of service activities are not new to child welfare (CW), over the last 30 years, federal legislation, including the Adoption Assistance and Child Welfare Act (P.L. 96-272) and the Adoption and Safe Families Act (P.L. 105-89), has promoted increased documentation in CW. Consequently, administrative CW data has proliferated and administrative data systems (ADS) have made these data more accessible to researchers.

To date, the majority of studies using administrative CW data have focused on the quantitative categorical data stored in ADS (see Conn et al., 2013; Putnam-Hornstein & Needell, 2011). Quantitative data help researchers and CW administrators identify rates of reported and substantiated child maltreatment, detect corresponding risk factors, or categorize service responses. The mining of these data teaches us about the kinds of maltreatment, placements, and services children referred to CW systems experience; identifies the frequency of these experiences; and can be used to make predictions about which children will return home and which will remain in care. However, these quantitative data tell us little about how CW workers define maltreatment, why children referred to CW systems are placed in specific settings, or how children and families engage in services. These latter questions are better answered through the mining and analysis of qualitative data stored in ADS.

Qualitative Data-Mining (QDM), the mining of the narrative text contained in documents stored in ADS (e.g., risk assessments, investigative narratives, court reports, and contact notes), provides CW researchers with a unique opportunity to use existing data to examine CW practice (Epstein, 2002, 2009). Use of QDM to improve CW has received limited attention (Epstein, 2002; Tice, 1998), as few CW studies have focused on the qualitative data stored in CW ADS or described how qualitative data is used by CW researchers (for exceptions see Coohey, 2007; Cordero, 2004; Cross, Koh, Rolock, & Eblen-Manning, 2013; Henry, 2014). This paper seeks to fill this gap by describing how
researchers can use QDM techniques to create rich databases for qualitative CW research and answer unique questions about CW clients and practice. In a seven-step guide, the paper summarizes QDM strategies and methods, and reports on the work of the Child Welfare Qualitative Data-Mining (CWQDM) Project to illustrate these methods and strategies. The paper concludes with a discussion of how QDM can be used to enhance CW practice, research, and education.

Project Background

The CWQDM Project developed in the context of a longstanding practice-research partnership between a university-based research center and a regional social services consortium involving the directors of 11 county social service agencies, the deans and directors of four graduate social work programs, and executive staff representing a local foundation (Austin et al., 1999). The CWQDM Project was designed in response to agency interests in developing their capacity to engage in QDM in CW. One county agency agreed to participate as the pilot site for the project. With our agency partner, the CWQDM Project sought to (1) create a CW database that could be used to examine CW practice, client needs, and emerging issues in the field; and (2) develop QDM techniques that could be replicated by CW agencies and research partners.

In the next section, we describe the specific actions and processes that we developed to carry out the CWQDM Project and, in seven steps, outline how CW researchers can use QDM to create retrospective databases for practice research. The description of each step includes a summary of major lessons learned, and the relevant literature is discussed throughout.

Step 1: Build (or Build on) a University–Agency Partnership

QDM requires a strong working relationship between university and agency partners. Trust, commitment, engaged leadership, and expertise are fundamental requirements, given the sensitivity and complexity of
CW data and the substantial investment of time and labor required to complete a QDM project. Projects should be of mutual interest, relevant to current practice, and provide equivalent benefits to both parties.

The development and success of the CWQDM Project was enhanced by the trust fostered through the regional consortium, shared research interests, and the intersecting areas of expertise among agency and university staff. The agency had a strong research and evaluation unit with expertise in CW research, policy, and practice. Members of the university research team had significant experience conducting CW research and were familiar with the CW practice context from prior work in the field.

**Step 2: Identifying Mutual Goals and Developing Practice Research Questions**

University and agency partners must identify mutual goals and work together to agree upon practice research questions. Even when there is agreement on research and practice goals, the best way to achieve these goals may be contested or limited by university and agency resources. Further, the type of qualitative data available in ADS will shape the kinds of questions that can be answered and the ability of the group to meet their identified goals. Before the university and agency can proceed with their project, the agency must provide an overview of the types of data available and the university must help the agency to understand what types of questions can be answered with these data.

Agency partners typically want to explore ways to enhance practice; increase efficiency; and meet state and federal child, safety, and permanency goals (Austin et al., 1999). University partners typically share the agency’s goal of enhancing practice and improving performance outcomes, but will have the added goal of contributing to the child welfare knowledge base through presentations, publications, and academic instruction. It is essential that all goals are articulated and agreed upon at the start of the project to ensure that goals are met and the project maintains ongoing support from all parties.
Once mutual goals are identified, specific research questions must be developed and agreed upon. In QDM, as a form of practice research, these questions should emerge from the field. Agency partners are particularly well situated to identify research questions pertinent to the field in general and their agency specifically. University partners are particularly good at identifying what is already known, noting gaps in the literature, and helping agency partners to develop researchable questions.

As noted above, our project sought to build a CW database to examine CW practice. As an exploratory project, the research questions were framed broadly. While the agency identified specific questions to pursue, such as the alignment between presenting problems and case plan development, the agency was willing to “let these data speak to us.” The university identified research questions that were important to the field, but agreed that letting these data guide us would further our understanding of what QDM could teach us about CW practice.

**Step 3: Identifying Practice and Research Concerns**

Concerns may arise early on about agency needs related to allocation of resources, data security, confidentiality, and dissemination of research findings. These concerns must be addressed from the outset of the project.

*Balancing Practice and Research Needs*

Qualitative research in CW can be intrusive, requiring direct access to key informants for interviews, surveys, or observation at the agency or in the field. These methods can yield rich practice data, but they can also be disruptive to daily practice and divert practitioners from meeting client needs. Alternatively, QDM offers a means of examining CW in a manner less intrusive to daily practice (Epstein, 2009). QDM does require time from some CW personnel; however, this time can be fairly limited and is concentrated mostly at the start and finish of the projects. At the outset, administrators and staff will be called upon to orient university researchers
to the ADS, develop information sharing agreements between the agency and the university, provide a policy and practice context for the researchers, work with university researchers to develop research questions and goals, assist with sampling, and facilitate data access.

In this project, the burden on administrators and line staff was fairly limited, in large part because of the relationship established through the regional consortium. Members of the agency’s research unit served as natural partners. Their prior work with the consortium, coupled with their own interest and expertise in CW research, ensured that there would be adequate levels of support for the project. The familiarity of the university researchers with CW policies, practices, procedures, the regional practice context, CW acronyms and the agency’s ADS also reduced the burden on agency personnel.

Confidentiality and Data Security

QDM raises significant concerns about client and agency confidentiality. Qualitative data contained in CW data systems are highly sensitive, often detailing personal information about vulnerable populations. In addition, qualitative data capture the daily practice of CW staff and the difficult decisions they must make. While these data can highlight client strengths, resilience and progress, and skillful CW practice, these data also reveal client challenges and subject agencies and their practices to scrutiny. Consequently, agencies may have concerns about how researchers will (1) protect the confidentiality of clients and staff, neither of whom have consented to participating in the research; and (2) how researchers will protect the confidentiality of the agency. Given these concerns, the researcher must develop a research protocol that protects the confidentiality of all parties and ensures that data retrieved from the ADS is secure. These research protocols should be reviewed and approved by institutional review boards (IRB) at the university level and by the agency and/or local courts.

Before beginning this project, all confidentiality and data security protocols were reviewed and approved by the university’s IRB. Data security
measures and confidentiality protections are described below. Data protocols were also approved by the executive director and CW director at the agency, as well as the head of the agency's research unit. Members of the university research team completed the CITI training in ethics and additional ethics training at the agency before accessing data.

**Dissemination of Research**

Dissemination of findings must balance client and agency confidentiality interests with researchers’ desire to share findings with the larger field. While IRBs require researchers to protect the confidentiality of their subjects and to provide detailed plans about how confidentiality will be maintained, including when disseminating findings, the research and agency partners should develop their own plan about how knowledge generated from the collaboration will be shared. Issues surrounding data ownership and dissemination of findings should be addressed early in the research process in an explicit protocol or memorandum of understanding. For this project, the partners agreed that the agency would play an instrumental role in deciding which analyses to pursue, and that all substantive findings emerging from the project would be reviewed by the agency before publication. Agency and university researchers agreed that whenever possible, findings would be presented together and publications would be written collaboratively.

**Step 4: Identifying Qualitative Data Sources and Assessing Data**

Working with administrative CW data is often messy (Epstein, 2009). When data are collected specifically for research purposes, they are often organized in a logical manner; the questions that researchers want answered are asked, and data are recorded and stored in a way that facilitates analysis. In contrast, administrative CW data may be stored across multiple paper files, hand-written notes may be illegible, and
important pieces of data (e.g., key documents and demographic data) may be missing. Increasingly, however, CW agencies rely on ADS to document their daily practice and to organize client information (Courtney & Collins, 1994; English, Brandford, & Coghlan, 2000). Researchers can now access most records from agency computers or even offsite from university offices.

Technological advances make it easier for researchers to access administrative qualitative data and to securely export these data to the cloud for offsite analysis. However, because these data were not created for research, researchers must take time to assess the quality of potential data sources (e.g., CW documents). Working with the agency partner is essential; agency staff can identify the richest data in the ADS, and explain the purpose of the data, when these data were typically created in the life of a case, and which data were mandated and therefore likely to be present in all records.

Once the researcher has become familiar with the types of data available in the ADS, the researcher and agency should work together to map these data onto CW practice. While the specific components of CW practice differ across agencies, most agency practice follows a similar flow. Specifying which aspects of practice the researcher seeks to understand, and which data sources in the ADS correspond to those practices, will help the researcher to determine which data sources to review. In our mapping of these data (see Figure 1) we were able to work with our agency partners to identify key data sources in the ADS that captured relevant aspects of CW practice.

After completing the mapping of data sources, the researcher should assess the quality of each source. To facilitate this review, the agency must provide the researcher with access to a sample of electronic case records. In our experience, this initial review can take place over one to two days and does not require the removal or extraction of any data from the agency. It is helpful to have staff on call during the review who can answer questions about how to navigate the ADS, the purpose of different data sources within the ADS, and how these data sources have changed over time.
Figure 1. Mapping child welfare process and data sources

1 Adapted with permission from Reed & Karpilow, 2009
When reading through qualitative data, the researcher should ask the following: (1) Are data captured in narrative form (e.g., “the mother reported being struck with a closed fist”) or check-boxes (e.g., “domestic violence in home? – yes/no”)?; (2) Are there enough data to discern meaning?; (3) Does data contained in one data source, consistently show up in another (e.g., are narratives from the initial investigation copied into subsequent court reports)? If so, can fewer data sources be examined?; and (4) What do these data sources not tell us? That is, what else does the researcher need to know about the practice context in order to accurately interpret these data? During this initial review it is important to note the format of each data source. Whether these data are formatted as text or image may affect how data are later extracted, stored, and coded.

**Step 5: Secure Data Extraction, Storage and Database Creation**

Data can be analyzed at the agency, but onsite analysis can be disruptive to practice, taxing on limited agency resources, and may not be practical for the researcher. Instead, it may be easier to store and analyze data offsite. The extraction and migration of these data requires the development of data extraction and storage security protocols. Electronic data can be copied, encrypted, password-protected, and securely transported on external hard drives to universities for secure storage and analysis. Alternatively, newer analytical software programs, such as Dedoose©, allow electronic data to be exported to secure cloud-based servers that can be accessed by researchers later. These cloud-based servers typically incorporate several levels of physical and electronic security measures designed to protect data and garner IRB approval.

In an effort to minimize our impact on agency function, we extracted and exported data to a secure cloud-based server. After working with our agency partners to select a stratified random sample of case records from the ADS, we developed an extraction manual that guided our research team through the ADS, pointed them to specific data sources (identified
through our prior mapping), and provided detailed descriptions on how to extract and label each data source; this included the assignment of unique research identification numbers to each case record and file naming conventions for each data source. Systematic labeling of case records and data sources at time of extraction enables researchers to organize data for later analysis and ensures that (1) all data sources are tied to a specific case record and can be identified through a unique research identification number; (2) the types of data sources associated with each case record can be identified without the researcher having to review the narrative text; and (3) that data sources can be sorted chronologically to aid future analysis. For our project, file-naming conventions included a unique research identification number, the date the data source was generated in the ADS, and an abbreviation of the data source type.

Over the course of five days, our research team extracted over 1,500 data sources from the agency’s ADS. Later these qualitative data sources were linked to the quantitative data previously provided by the agency through the research identification number assigned to each record to create a CW database that could be used for mixed methods analyses. Just as ADS provide CW agencies with a means of organizing case records and corresponding documents for practice, our data extraction protocols and data linkages allowed us to create a CW database that organizes case records and corresponding data sources for practice research.

**Step 6: Generating Practice Knowledge: Analytical Strategies**

The first five steps created the foundation for a CW database that could be used to examine CW practice, client needs, and emerging issues in the field. Equally important was the development of analytical techniques to examine and understand these data. Researchers can choose from a range of analytic approaches (see Miles, Huberman, & Saldaña, 2013 for a description of different approaches to qualitative data analysis); the approach chosen should be driven by the type of questions the research team seeks to answer or the overarching goal of the project.
We adopted two primary data analysis strategies, *code-based analysis* and *case-based analysis*. Both strategies were employed simultaneously and both offered different insights to how QDM might be used to inform CW practice (see Figure 2).

**Code-Based Analysis**

Code-based analysis was employed to gain a better understanding of the types of phenomena that were captured in the CW database (see Figure 2). While Grounded Theorists suggest that researchers enter the field or approach the text, with “an open mind” and let problems and themes emerge through open-coding (Glaser, 1992, p. 23), Gilgun (2005, p. 42) argues that researchers and practitioners should not have to “forsake well-formulated conceptual models” when engaging in qualitative research.

**Figure 2. Knowledge for child welfare practice: What we can learn from QDM**
Instead, Gilgun (2005) suggests that researchers employ what she has coined, “deductive-qualitative analysis,” which allows researchers to develop preliminary deductive codes based on their pre-existing knowledge or conceptual models and then develop additional inductive codes through open-coding of data. Using this analytic approach, our research team developed a preliminary deductive codebook based on Belsky’s (1980) ecological framework for child maltreatment. This framework conceptualizes child maltreatment as a “social-psychological phenomenon” that is determined by forces at work in the individual, the family, the community, and the culture (Belsky, 1980, p. 320).

The research team piloted the preliminary codebook through review of the data sources for a sub-sample of case records. Consensus was reached among the group that the codes included in the codebook reflected phenomena observed in the data. During this initial review and subsequent reviews, new codes were identified through inductive open-coding. These codes captured CW services and processes, actions and inactions, unique social problems (e.g., child exposure to domestic violence) and information about the experiences of poorly understood populations (e.g., commercially sexually exploited youth) (see Figure 2). During a four-month period the codebook was revised through an iterative process that generated over 65 unique codes nested under twelve broad themes.

The coding of these data mapped the types of phenomena captured in the CW database. Each code can be used as an entry point into the database for future in-depth analyses. For example, in our coding we identified and labeled all narrative data related to the experiences of kin caregivers. Additional analyses could examine how CW staff work with kin caregivers to support children and birth families and the role this cooperation plays in establishing permanency for youth. These types of code-based analyses enable us to identify patterns that are common within a subset or across all CW cases. Building on these patterns, we can develop hypotheses for later testing. In addition, where categorical quantitative data are not available, code-based analyses can be used to estimate prevalence of a specific phenomenon. This is particularly useful for capturing emerging social problems.
Case-Based Analysis

While coding these data, our research team simultaneously engaged in case-based analysis. We used a within-case analysis approach, the goal of which is to “describe, understand, and explain what has happened in a single, bounded context” (i.e., the case or site) (Miles et al., 2013, p. 100). We used a pre-structured case outline to summarize key aspects of each case in the database. The close reading required for our code-based analysis ensured that case summaries captured more consistent and accurate data across all cases. We began by reviewing all qualitative data sources associated with a single case. From this review we then developed a summary of the case that included: a description of child and family characteristics; an overview of key stakeholders, their relationship to the child, and initial and emerging problems; and a summary of case events. In addition, the summary included a detailed timeline of significant events and the reviewer’s critical reflections on the case.

Cross-case analysis (Miles et al., 2013) can also be employed to explore the inherent complexity of CW practice and the common and divergent trajectories of cases over time (see Figure 2). Cross-case analyses “increase generalizability” and ensure that “events and processes in one well-described setting are not wholly idiosyncratic” (Miles et al., 2013, p. 101). Analyses of similarities and differences within and across cases can help CW researchers to better understand client needs and how individual CW practice is shaped by the local, state and federal context.

Analysis and Review

After a multi-day training on the qualitative analytic software, codebook structure and application, and case summaries, each member of the research team was assigned a unique set of CW cases to code and summarize. Initial coding and summaries were carried out in a single office space, so that the research leaders could respond to coding and summary questions as they arose. Throughout the duration of the project each member of the research team was encouraged to conduct their analysis in a space shared with other team members to facilitate
ongoing discussion of code application and enhance intercoder reliability (MacQueen, McLellan, Kay, & Milsten, 1998). Throughout the analytic process the project leaders reviewed and edited case summaries for completeness and clarity.

Qualitative methodologists suggest that researchers check their preliminary findings with key informants during the research process (Guba & Lincoln, 1989). Throughout our project we discussed our research process and analytical frameworks with members of the regional consortium and agency partners. Agency research staff reviewed preliminary codebooks, provided policy insights, clarified the meaning of local terminology and acronyms, and encouraged us to further analyze themes that were of particular importance to agency practice. We provided interim reports and presentations to consortium members and agency staff to gain their feedback on the patterns and themes we had identified, the case summaries we had generated, and their relevance and utility for practice. This “member-checking” (Guba & Lincoln, 1989) served to validate our research, and also helped us to identify which practice research directions we should pursue given the available data in the CW database and the CW needs in the region.

**Step 7: Generating Practice Knowledge: Dissemination of Findings**

Dissemination of the knowledge generated through QDM can be accomplished through a range of approaches. Dissemination to CW agencies may involve internal reports, conversations with CW administrators or presentations to staff. Dissemination to the field includes academic instruction, government reports, conference presentations, and publication in academic or professional journals.

CW databases and the qualitative analyses that emerge from these databases can also serve as valuable training tools for CW staff. CW staff rarely have the opportunity to read an entire case record; notable exceptions include rare critical incident reviews and periodic case record review processes conducted for compliance purposes (Carnochan,
Samples, Lawson, & Austin, 2013; Douglas & McCarthy, 2011; The Children’s Services Outcomes and Accountability Bureau & The Office of Child Abuse Prevention, 2014). Generation of case summaries and cross-case analyses from CW databases allow CW staff to examine how other staff meet client needs and how cases unfold over time. Review of these summaries can help staff to identify promising practices and areas for improvement. In addition, researchers can make CW databases available to partner agencies or other CW agencies so they can conduct their own analyses or use the data for training purpose.

QDM and CW database creation can also be used to educate future CW staff or others in the helping professions. For example, by recruiting and training MSW student research assistants to help with the extraction, organization, and analysis of qualitative CW data, our project served to familiarize future social workers with QDM, ADS, client needs, promising practices, and the complexity of CW. Many students used the database and case summaries generated by the CWQDM Project to carry out their own master’s-level research projects (see Figure 2).

Discussion

QDM offers CW researchers and agencies a relatively non-intrusive means of examining CW services and generates new knowledge about CW clients and practice that cannot be gleaned from quantitative data alone. Unlike other qualitative methods, such as surveys, interviews, and participant observation, which only capture data from research participants who opt-in, QDM offers researchers the potential to capture the practice experiences of all CW staff and clients (Epstein, 2009). With respect to clients, QDM allows us to gain a more nuanced understanding of CW populations, particularly the prevalence of new and known social problems; poorly understood sub-populations; and complex client needs. Regarding practice, QDM provides new insights into promising practices, case trajectories, and case planning; how CW agencies and staff define and respond to parental acts and omissions; and how agencies work to enhance child safety, permanency, and well-being in daily practice (see Figure 2).
Despite its utility, QDM also raises critical challenges with respect to the complexity, trustworthiness, volume, and sensitivity of the qualitative data contained in ADS. The complexity and trustworthiness of these data relate in part to the presence and absence of multiple perspectives captured in data sources over time, as well as missing data. Data for each case are entered into ADS by multiple CW staff, resulting in a case record that reflects multiple perspectives. When these perspectives identify similar client strengths and needs or effective interventions, the trustworthiness of the data may be enhanced (Creswell & Clark, 2006). However, all data contained in ADS are filtered through the CW staff perspective. Data stored in ADS describe maltreating behaviors, client needs and wants, and services rendered, yet all of these data, even when recorded in the client’s voice, reflect what CW staff saw, heard, or were told and may not accurately represent events or the perspectives of those outside the agency. Similar narratives by CW staff may not point to truth, but instead may demonstrate how client identities are similarly constructed by staff (Swift, 1995; Tice, 1998). In addition, important data may not be recorded in the ADS; as a result, critical knowledge about practice and clients may be missing.

The complexity of these data are intensified by their sheer volume. Data often capture CW practice across years; narrative text recorded in one data source is often copied into another; and data are often not recorded by staff in-vivo, making it difficult for the researcher to piece together a linear account. Reliably coding these data presents challenges for research teams. While codebooks, consensus-based coding, qualitative analytic software (e.g., Dedoose©, Atlas TI©) and automated text analysis (e.g., Python©) help to improve intercoder reliability (Miles et al., 2013), the volume of these data poses significant challenges.

Data contained in ADS are highly sensitive, and use of these data for research purposes present threats to confidentiality. While data can highlight promising CW practices, the practices and experiences recorded in ADS could place agencies at risk of community censure, legal liability, or fiscal sanctions and could serve to stigmatize both CW workers and clients. Strategies to protect confidentiality through
de-identification include algorithmic methods used in conjunction with human review to locate and modify identifying data.

Despite these challenges, QDM offers substantial knowledge building opportunities for CW agencies, researchers, and future CW staff. Through engagement in QDM, CW staff becomes both producers and consumers of practice knowledge (see Figure 2). Participation in this process and familiarity with these data may decrease staff resistance to use of data reported in some studies, and may enhance the importance staff give to documentation of practice (DeFraia, 2015; Hutson & Lichtiger, 2002). QDM, when done in partnership with CW agencies, provides researchers with an opportunity to utilize existing data to enhance agency practice and share new CW knowledge with the larger field. QDM also offers future CW staff (i.e., students) an opportunity to better understand and generate CW knowledge. Engagement in QDM and qualitative analysis provide both current and future CW staff new skills to make meaning of complex data. Finally, QDM and the CW databases that can be constructed with these data offer CW researchers and agencies the ability to look beyond the quantitative data that describe basic caseload characteristics and performance outcomes, and instead begin to examine how CW agencies work to meet their clients’ complex needs in daily practice.

References


