Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

This document was prepared to accompany other resources on evaluation of Quality Rating and Improvement Systems (QRIS) and other quality improvement initiatives developed by the Quality Initiatives Research and Evaluation Consortium (INQUIRE).

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre/index.html.

Overview

As a centerpiece of state early care and education (ECE) activities, Quality Rating and Improvement Systems (QRIS) serve as an example of the how an effective ECE data system can support planning, operations, service delivery, monitoring and evaluation. Intentional and rigorous data management practices are essential for data gathered exclusively for the QRIS (such as program observation scores), as well as for external data accessed by the QRIS (such as workforce registry data). Implementing strong ECE data governance and management practices will ensure the quality of QRIS data and thus the integrity of the QRIS itself. Incomplete, inaccurate, or unreliable data can introduce errors in the ratings that can threaten the credibility of the QRIS and have negative consequences for ECE and school-age care (ECE-SAC) programs through skewed reimbursement rates and inaccurate marketing tied to incorrect ratings.

The purpose of this brief is to illustrate the need for and benefits of building strong ECE data governance structures and implementing system-wide data management policies and practices, using the example of QRIS.

Key informants interviewed for the brief articulated challenges of data systems that support QRIS:

■ States use data from databases governed and administered by multiple agencies and organizations.
■ States typically lack a governance framework for ECE data systems and management.
■ Databases accessed by the QRIS have been designed for different purposes. As a result, data coverage and availability are limited, and linking data presents challenges.
■ Data practices often do not support the production of high quality data.

Many states are designing or redesigning their QRIS data systems and are looking for models and guidance to inform their process and to address the challenges outlined above. The brief provides examples of four systems and their characteristics: unlinked databases or point solutions; linked data system based on customized interfaces between databases; federated, shared data system; and comprehensive, integrated data system.

As states build QRIS using existing ECE data, they also address a number of issues, regardless of the ECE data system strategy selected. Creating a clear and effective governance structure for an ECE data system is an essential component of system-building. The brief recommends that states:

■ Identify a governance body for the ECE data system;
■ Develop partnership and data-sharing agreements across all programs that provide ECE data;
■ Develop documentation for databases in the ECE data system;
■ Have a policy regarding database updates;
■ Develop common data standards;
■ Determine unique identifiers for children, workforce, and facilities;
■ Train and cultivate data management staff;
■ Ensure that a paperless system is also a valid and accurate system; and
■ Establish consistent security and back-up policies.

Data governance and management are terms that are new to many in the ECE field. This brief documents the challenges of data management in ECE data systems and provides concrete guidance for creating the infrastructure and governance that can address these challenges.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 1

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Introduction

Building an integrated, improvement-driven early care and education (ECE) data system is increasingly viewed as a critical foundation for strengthening the quality of early care and education and school-age care (ECE-SAC) programs and improving overall ECE system functioning. An effective ECE data system integrates child-, family-, program-, and workforce-level data in a manner that supports decision-making and continuous system improvement. A typical ECE data system includes data from a variety of programs and activities such as child care licensing, state pre-kindergarten programs, Head Start programs, special education, home visitation, quality improvement initiatives, and workforce/professional development registries.

As a centerpiece of state ECE activities, Quality Rating and Improvement Systems (QRIS) serve as a salient example of the how an effective ECE data

system can support planning, operations, service delivery, monitoring and evaluation. A QRIS and the ratings it produces are grounded in data about facilities, the workforce, and the children who are served by them. The ability of a QRIS to strengthen ECE-SAC program quality is directly related to the quality of the data on which it is built. Producing the ratings involves collection, management, and analysis of data from a variety of sources about ECE-SAC facilities and the people who work in them. Intentional and rigorous data management practices are essential for data gathered exclusively for the QRIS (such as program observation scores) as well as for external data accessed by the QRIS (such as workforce registry data). Implementing strong ECE data governance and management practices will ensure the quality of QRIS data and thus the integrity of the QRIS itself. Incomplete, inaccurate, or unreliable data can introduce errors in the ratings

Best Practices in Data Governance and Management for Early Care and
Education: Supporting Effective Quality Rating and Improvement Systems

Page 2

that can threaten the credibility of the QRIS and have negative consequences for ECE-SAC programs through reduced reimbursement rates and negative marketing tied to incorrect ratings.

The purpose of this brief is to illustrate the need for and benefits of building strong ECE data governance structures and implementing system-wide data management policies and practices, using the example of QRIS.

The brief first outlines the need for strong ECE data governance and policy by describing existing QRIS data systems and the common challenges to data coordination and integrity in these data systems. The brief then provides guidance on best practices related to data governance and the development of integrated data systems that can support QRIS implementation, monitoring and evaluation. A companion brief (Friese, Tout & Kirby, 2014) describes best practices related specifically to QRIS data management processes that ensure data quality (i.e., designing data systems that reduce errors, the treatment of historical data, and the necessity of strong system documentation).

This brief adds to a growing body of work on ECE data systems and elements of an effective QRIS, developed by the Quality Initiatives Research and Evaluation Consortium – INQUIRE – a project of the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families, U.S. Department of Health and Human Services. INQUIRE provides research-based guidance on developing and evaluating QRIS and other quality improvement initiatives; it also serves as a learning community for researchers and state partners engaged in QRIS evaluation efforts. This paper builds on INQUIRE products such as the *INQUIRE Data Toolkit* (2013), as well as products from the Quality Rating System Assessment project funded by OPRE from 2008-2011 (see the Reference section for citations and links to these resources). The *INQUIRE Data Toolkit* is especially important to reference in combination with the information in this brief because it provides an overview of common data elements (with data definitions and formats), and recommendations for analysis using the data elements to address policy questions.

> Implementing strong ECE data governance and management practices will ensure the quality of QRIS data and thus the integrity of the QRIS itself.

Approach

The development of integrated ECE data systems (or Early Childhood Integrated Data Systems -ECIDS) is growing due in part to funding from the Statewide Longitudinal Data Systems Grant Program and the Race to the Top – Early Learning Challenge Grant and through work by the Early Childhood Data Collaborative.[1] This brief was developed to focus in particular on the needs of QRIS administrators who are frequently on the front lines of data governance and data management issues in the ECE system. For more information about Early Childhood Integrated Data Systems (ECIDS), please see resources developed by the National Center for Education Statistics (2014).

In developing a guidance document, we consulted with state and local QRIS data managers and administrators who could provide the most relevant examples of the current practices and related challenges in QRIS data systems and their connections to broader ECE data systems. In addition, we realized we could benefit from the experience of informaticists who oversee the coordination of disparate statewide databases, primarily in the public health sector. The discipline known as informatics focuses

1 The Early Childhood Data Collaborative (ECDC) provides tools and resources to states to aid in the development of coordinated data systems with the ultimate goals of improving program and workforce quality, increasing access to high-quality early childhood programs, and improving child outcomes.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 3

on the content, representation, capture, and use of information to support workflow and promote best practices. Thus, our approach to developing the content of this brief was to conduct interviews with state QRIS data managers and a public health informaticist to identify best practices in data governance and strategies to develop an integrated ECE data system that can support an effective QRIS. The authors also engaged national research experts to review and provide input on drafts of this brief.

Data Collection Details

With the help of the Quality Rating System Assessment Project,[2] the authors identified states to interview about their QRIS data management protocols and procedures. The project team also helped identify representatives who could contribute knowledge about that state's QRIS data system. Eight states and/ or localities with a range of QRIS experience were selected: Miami-Dade County, Florida; Indiana; Maine; Minnesota; North Carolina; Ohio; Oregon, and Virginia.[3] QRIS data managers and, in some instances, state professional development registry and child care resource and referral personnel were interviewed. Interviewees were asked a unique set of questions depending on their role in their state's QRIS. For example, some individuals were asked questions based on their familiarity with the collection of observational data from child care programs while others were asked about the professional development registry data they worked with regularly.

The length of interviews ranged from 30 to 60 minutes. Participants were asked questions about how QRIS data in their state is stored and managed, including: a) the number and types of databases used within their QRIS; b) whether there are written principles, policies, or procedures for collection, entry, and management of QRIS data; c) how identification numbers (IDs) are assigned and used; and d) when applicable, what processes/steps are used in managing the quality of the data (e.g., random checks, testing for errors in data, existence of a codebook for the QRIS data, training for individuals in the data management process, and ownership and responsibility of the QRIS data)(see Appendix A for details).

In addition to interviews with state QRIS data managers, a series of interviews was conducted with an Oregon Health Authority informaticist. These interviews provided information to identify and translate general principles from the field of Informatics to be relevant to developers and administrators of QRIS. Data from these interviews were reviewed and analyzed by the authors to identify key themes related to the current practices, challenges, and opportunities for increasing coordination in QRIS data systems. These interviews also provided a basis from which to recommend best practices for building broader ECE data systems.

---

2 The Quality Rating System Assessment Project was a project conducted from 2008 to 2011 that had five goals: 1) to develop a compendium describing states' QRS; 2) to complete in-depth case studies of a few QRSs; 3) to conduct secondary analyses using QRS evaluation data; 4) to create a paper synthesizing the case studies and secondary analyses; and 5) to develop a QRS best practices toolkit for states. The project was funded through the Office of Planning, Research and Evaluation in the Administration for Children and Families.
3 Two Quality Rating System Assessment Project products are particularly relevant for the information described in this brief. First, Caronongan and colleagues (2011) conducted an in-depth study of five state QRIS that included two of the state/localities interviewed for this paper (Indiana and Miami-Dade). Second, Malone and colleagues (2011) conducted secondary data analyses with QRIS data from three states and gained insights into the challenges of working with QRIS data. Malone is also a member of the INQUIRE Data Workgroup and added insights from that work to this brief.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 4

A QRIS Lens on Common Characteristics, Challenges, and Opportunities for Improving ECE Data Systems

Though each state has unique features, common characteristics of data systems supporting QRIS emerged from the interviews with eight state or local QRIS managers. This section describes the key characteristics and the challenges and opportunities to improving data governance and coordination that are associated with each.

1. **States use data from databases governed and administered by multiple agencies and organizations.**

QRIS ratings are generated by drawing on data from a variety of sources. For example, states commonly use data from the child care licensing system, state prekindergarten program, Head Start programs, and workforce/professional development registry to inform the designation of QRIS ratings. Subsidy data is used less often to inform ratings but may be used to determine the number of children receiving subsidies enrolled in individual facilities (which in turn may determine subsidy reimbursement rates and bonuses in some states). Databases to document the provision of technical assistance and other investments in ECE-SAC programs and practitioners (e.g., facility incentives, WAGE$, T.E.A.C.H. scholarships) may also be in place.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 5

In some cases, states have built QRIS data systems on existing databases to avoid costly duplication and the errors that come with capturing the same information on the same organizations and/or individuals in different databases. In addition, states may create new databases to manage data collection on key elements that are not available from existing data sources, such as QRIS application and enrollment data and program observations (e.g., data from the Environment Rating Scales, the Classroom Assessment Scoring System, and/or state-developed instruments), that are used to establish a rating.

2. **States typically lack a governance framework for ECE data systems and management.**

A critical challenge for QRIS is that many states do not have established authorities that govern and manage the policies and practices of data in the ECE system. For example, even when a single state agency controls funding for all of the programs whose databases are used in QRIS, typically each database has its own administrative structure with authority for how that database is managed. Additionally, many states do not have well-articulated processes of how QRIS data are to be managed. In many instances, individuals enter data as a part of their many duties without clear guidelines on how data should be entered, merged, checked, or stored. States that have individuals assigned to managing the data are more likely to have up-to-date procedures and documentations that ensure high-quality data for their QRIS, but without an ECE data governance system these procedures will vary depending on who administers the source databases. Without comprehensive common data standards and well-articulated data management practices shared by all ECE organizations, agencies, and programs that contribute data to QRIS, the system may use inaccurate data. Numerous policies and practices that affect data quality may vary across the databases linked for QRIS; for instance, there may be difference in how data are linked and in policies on data definitions, confidentiality, privacy, security, and who may access data.

3. **Databases accessed by the QRIS have been designed for different purposes. As a result, data coverage and availability are limited, and linking data presents challenges.**

Licensing databases typically include data on ECE-SAC facilities, while professional development registry databases typically contain data on demographics, qualifications, credentials and training of the ECE-SAC workforce. When a state uses existing databases for its QRIS or other purposes, it is important to understand the purpose as well as the ownership of databases from other agencies or other state or local systems, and the related challenges.

Additionally, in some administrative databases, data may only be collected for data elements that are essential for the work of that particular agency or organization, such as providing funds or awarding a license. Data elements that are discretionary and have no implication for funding or licensing may be left blank by individuals collecting and entering data, resulting in missing data for a number of elements that may be important to establishing a QRIS rating.

Another challenge is that database practices, such as overriding variable values when updating records, impede data sharing and other uses. The structure of the database files often does not accommodate longitudinal data collection that would enable the detection of change over time. A final issue encountered when accessing available data is that each agency and organization has its own method of data representation and record identification, making it difficult to link at the individual (e.g., child, ECE-SAC workforce member) or facility (e.g., child care center, family child care home) level, and limiting what can be known about the individual or organization.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 6

4. **Data practices often do not support the production of high-quality data.**

Both data collection and data linkage affect the quality of data. Data integrity begins in data collection processes, which may also include the linking of data. In some states, data are collected by multiple agencies with no common policies or documentation about how data are entered and stored.

States, increasingly, are using web-based systems for data entry and storage. Electronic linking is becoming more common, but some states move data manually from one database to another, either through manual data entry or through merging data files. Most states do not have a technical manual describing how QRIS data are merged, or detailing the various procedures in place to ensure high accuracy and low errors.

5. **States are designing or redesigning their QRIS data systems and are looking for models and guidance.**

While some states are building their first QRIS data system, other states are redesigning their QRIS data system by improving automation or adopting a web-based platform. Many states are becoming more web-based as a way to facilitate linkages between and among databases and to be financially sustainable while ensuring high-quality data for their QRIS. As their systems are changing, states are finding challenges in determining the optimal data system for their state. This period of rapid change is an ideal time to support states in building QRIS data management capacity. Yet resources and time are limited. Design and redesign of data systems may be done under tight timeframes and with limited resources for engaging consultants or staff with expertise in data science, governance, and management. Requests for support within state agencies might not be prioritized.

Thus, existing QRIS data systems and efforts to design new data systems face multiple challenges. The urgency surrounding the establishment of ECE integrated data systems with sound data governance and data management practices creates an opportunity for reflection and intentional planning to address challenges and to put practices in place that can sustain the production of high-quality data across the ECE system.

In the next sections of the brief, we provide practical guidance on developing integrated ECE data systems that can support QRIS and other ECE system initiatives.

Basic Options for Developing an Integrated ECE Data System

When a state decides to build an ECE data system, it has four basic options. The four options are described in Figure 1, with the top three describing ever-more integrated systems. The bottom of the figure represents databases that are not linked and not capable of supporting efficient data sharing and integration. The next layers of the figure display greater and greater coordination and – at the top – full integration of the data into a shared data system. An integrated data system offers standards-based data that is entered only once and then used wherever needed according to permissions of the user. Not linking or coordinating existing databases is likely to result in duplicative efforts and substantial error as the same data are entered into multiple databases and manually or electronically brought together at the individual level.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 7

Figure 1. Early childhood development data system options

System Type	Data Quality	Data Availability	Cross Agency Workflow	Data Governance	Description
Comprehensive, integrated data system	Highest	Highest	Most efficient	Most efficient	Data elements needed for QRIS and other purposes are addressed in a single integrated data model. Data silos are eliminated which also eliminates redundant data entry and workflow. Data are represented according to standards, so quality is high. Data are available for all needs at the time it is entered. The governance process includes stakeholders from all participating agencies.
Federated, shared data system					Data elements needed for QRIS and other purposes are extracted from databases, mapped to standards, linked to master identifiers and stored in shared repository. Individual databases are in place, but shared data is used for cross-program functions. Data are available for cross-agency needs after processing daily or according to a regular schedule. Cross-agency governance is required for shared data, but individual databases may retain their own governance process.
Linked data system based on customized interfaces between databases					Databases are linked one by one as needed. All functionality resides in the individual database. Data are combined for reporting or other needs. Data are not based on standards, so quality may be poor. Interfaces are designed ad hoc and require ongoing maintenance. Data are available only to programs with interfaces and may require processing prior to availability. Governance occurs at the agency level, and sharing is addressed on a case-by-case basis.
Unlinked databases or point solutions	Lowest	Lowest	Least efficient	Least efficient	No data or functionality are shared across programs creating redundant data entry and workflow burden. Databases are created for single purposes without uniform standards. Sharing data takes time and resources, and data quality may be poor. Data are available only to the source agency. Governance occurs at the agency level for each database.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 8

States must weigh a number of factors when deciding which option is the best fit for the short term and long term. For a QRIS, data quality is essential, as the credibility of a QRIS rests on accurate and reliable data for creating ratings. Databases also have to meet the programmatic needs of organizations, such as licensing. Even with cost savings over time, change to source databases will be disruptive to the organizations that rely on them to manage their programs. Impact on functioning of the multiple programs whose data are used is an issue that needs to be considered. Weighing the expense of the different options is complicated. For example, a comprehensive integrated system may be a cost-effective option in the long term but it entails the greatest change in how programs operate. It is worth noting that moving from databases that are managed independently by different programs into a comprehensive integrated data system mirrors the transformation from independent programs operated by multiple organizations into an ECE system. Integrating databases may well be essential to reaching the broader goal of an integrated early care and education system.

Options will look different in states that have established data warehouses that include child care and early education databases. In those cases it may be that work will focus on enabling the central repository of the data warehouse to link data from multiple ECE databases at the child, facility, or workforce levels. If not all ECE databases are included in the data warehouse, strategies for coordinating all needed databases will need to be created.

Determining the best data system option is not easy. If a state has to choose a less-than-optimal option, they will need help determining how to address some of its associated limitations. States may want to access experts from the field of informatics to help them understand their options and determine which best fits their current and anticipated needs (see text box above).

Best Practices for Creating Strong ECE Data Governance and Management Structures

As states build QRIS using existing ECE data, they must address a number of considerations, regardless of the ECE data system strategy selected. Creating a clear and effective governance structure for a state's ECE data system is an essential component of system-building. In this section, we share guidance for creating and managing a state's ECE data governance and management plan. It is important to note that the work of building an ECE data system will most likely be occurring in the context of broader work on a full early childhood data system (going beyond early care and education to include health and family support, for example) or other bigger data system development activities. Activities for strengthening ECE data should be coordinated with these efforts.

Identify a Governance Body for the ECE Data System. A first task in setting up effective governance is to identify a body that will manage the process. The Early Childhood Data Collaborative[4] recommends that states have a governance body that "establishes the vision, goals and strategic plan for building, linking, and using data to support continuous improvement ... and guiding data collection, access and use" (2010).[5]

---

4 According to its website, "The Early Childhood Data Collaborative (ECDC) supports state policymakers' development and use of coordinated state early care and education (ECE) data systems to improve the quality of ECE programs and the workforce, increase access to high-quality ECE programs, and ultimately improve child outcomes. The ECDC will provide tools and resources to encourage state policy change and provide a national forum to support the development and use of coordinated state ECE data systems."

5 Early Childhood Data Collaborative, www.ecedata.org. A number of organizations and resources exist that can be consulted to support state work on data governance. The Data Management Association (DAMA) is one possible source. The Data Governance Institute (DGI) and the governance model created by IBM are other possible sources to consider.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 9

Develop multiple levels for the governance body to ensure that implementation is supported. Because of the different responsibilities required in data governance, multiple levels of governance may be needed. The following are examples of governance levels adapted for an ECE data system.[6]

■ An *Executive Council* sets the overall vision, mission and strategic goals of data governance. The vision should align with the overall vision, mission and goals for the ECE system as a whole and for the QRIS as a component of that system. The Executive Council is also responsible for obtaining needed funding and resources.

■ The *Strategic Committee* develops the high-level task plan to achieve the goals established by the executive council. Strategic Committees could be developed, for example, to support the QRIS and/or professional development data components.

■ The *Tactical Group* develops short-term goals and tasks to implement the high-level plan mandated by the strategic committee. It includes data stewards and subject matter experts (for example, QRIS data managers) as members.

■ *Partners and other stakeholders* should be engaged regularly to provide ideas and feedback to the formal management organization for data governance.

If successful, the governance body and the structures that are put in place can address the challenges of "turf, trust, technical issues and time" that characterize interactions about data (Data Quality Campaign, 2012). According to the Data Quality Campaign,[7] "turf" is addressed by bringing multiple agencies and leaders to a common table and offering a forum for airing issues and creating a shared responsibility for the data system. "Trust" is addressed by developing agreed-upon standards for collecting and using data and ensuring data quality and security in the system. "Technical issues" are addressed by putting policies in place that can resolve differences across data owners in technical standards or processes. "Time" is addressed by creating clear roles and responsibilities that increase the efficiency of work in the system.

Carefully plan the membership of the governance body. The governance body should include a director who is responsible for the work and who can manage and monitor ongoing operations. Other vital members include agency directors, agency staff responsible for information technology, data architects (programmers), data stewards (managers), subject matter experts (program staff), data analysts, and data consumers or users. In particular, research and evaluation staff and/or consultants should be included to ensure that decisions made about data are aligned with the needs of reporting and evaluation.

Develop Partnership and Data-Sharing Agreements Across All Programs that Provide ECE Data. The governance body serves as the infrastructure for developing partnerships with agencies to decide how data systems can be interfaced and the type of information that is needed to meet the goals of the system. An essential task is to develop a standard memorandum of agreement on which all data sharing agreements are based. Note that this step is not necessary once states have reached a truly integrated ECE data system.

---

6 The model is adapted from work produced for managers of Federal Student Aid programs for students (2007, p. 8).

7 The Data Quality Campaign (DQC) is a nonprofit, nonpartisan, national advocacy organization based in Washington, DC. Launched in 2005 by 10 founding partners, DQC now leads a partnership of nearly 100 organizations committed to realizing the vision of an education system in which all stakeholders—from parents to policymakers—are empowered with high-quality data from the early childhood, K–12, postsecondary, and workforce systems to make decisions that ensure every student graduates high school prepared for success in college and the workplace. To achieve this vision, DQC supports state policymakers and other key leaders in promoting the development and effective use of statewide longitudinal data systems. The Data Quality Campaign, in partnership with the Early Childhood Data Collaborative, surveyed 48 states and the District of Columbia in fall 2010 to track state progress toward implementing the 10 Fundamentals of Coordinated State ECE Data Systems.

Best Practices in Data Governance and Management for Early Care and
Education: Supporting Effective Quality Rating and Improvement Systems

Page 10

Develop Documentation for Databases in the ECE Data System. Creation of comprehensive data management documentation is needed for the databases used in the ECE data system. These documents should have information about all the variables in the data system, including how summary variables are generated. If standard procedures are not yet incorporated across the system, the documentation should also describe how data from other resources are included in the database and the processes and checks involved to limit errors. Standard quality assurance practices include double data entry with verification, range checks built into all fields when possible, validity checks across fields – especially involving IDs, and a stated process for addressing potential errors identified with the checks. This documentation should be electronic and include a history of changes in data management, coding and programming.

Have a Policy Regarding Database Updates. A common practice is to override a value when updating a database. It is critical that a new dataset or variable be created when updates are made and checks run on that new dataset, and that changes are documented. Furthermore, changes to data should be programmed rather than done manually to minimize errors. Strategies have been designed to enable administrative databases to provide longitudinal data. An example of such a strategy is the insertion of date variables attached to variables that change over time (e.g., a level on a registry or a rating).

Develop Common Data Standards. Common Data Standards have three key elements:

1. Set of data elements: a minimum data set (MDS) (e.g., all partners agree to collect each element in the MDS although partners can collect data elements in addition to the minimum, core set)

2. Syntactic representation: agreement on how each element is represented (e.g., is it a number, and if so how many decimals; is it text, and if so how many characters; is it encoded, and if so which code set will be used?)

3. Semantics: What does it mean? What question is asked of the respondent? (For example, it is difficult to describe teacher education because of differences in education questions asked and in ways education levels are measured. If different questions are asked about the amount of education attained, one cannot assume that the responses have the same meaning.)

In addition to common data standards, data management standards should be developed that include agreed upon protocols for dealing with security, privacy, access, and change.

Determine Unique Identifiers for Children, Workforce, and Facilities. Avoiding duplicate records and being able to link with other systems requires a unique ID number for each individual (i.e., child and workforce) and facility. For children, states sometimes adopt the State Student Identifier (SSID) used in K-12 or the unique number assigned by the state's health system. Licensing IDs are typically used for facilities, which raises issues for inclusion of non-licensed programs that have to be resolved if the state wants all early learning programs to be included in the QRIS. The role of unique ids in an ECE data system is important because it helps connect all the data about an entity and provides a basis for analysis of what is happening for programs, children and the workforce.

Train and Cultivate Data Management Staff. It is imperative to ensure that all data entry staff and managers are trained on data entry policies and provided with refresher training. This would also require a data management protocol policy book that is user-friendly but also detailed enough to prevent errors. Additionally, it is valuable to have an individual and/or system for monitoring, training, and documenting all aspects of the data management process and protocol.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 11

Ensure Paperless System is also a Valid and Accurate System. A goal of data system development is to eliminate paper and have only electronic records when possible. States are at various stages of reaching this goal. In the interim, a goal may be to minimize the amount of paper that is used and determine how best to validate the information.

Take Advantage of Technological Advances. Technological advances make possible the integration of large amounts of data from multiple systems—especially when internal checks can be programmed to limit discrepancies in data entry—as well as ensure high-level security. Use of pull-down menus rather than reentry of values such as IDs reduces data entry errors.

Establish Consistent Security and Back-up Policies. The data system should be securely stored and backed up regularly. Access to data should be determined and documented with access and level of access depending on role of each individual in data management and QRIS (e.g., full access to administrators, read-only access for researchers). Recommended practice is to conduct nightly backups that include all files that changed during the day and weekly backups of all files. It is essential that a regular backup schedule be established, and that a copy of recent backups is stored off-site as well as on-site (Burchinal and Neebe, 2006).

Conclusion

Data governance and management are terms that are new to many in the ECE field. QRIS administrators are managing the implementation challenges of a system that brings together multiple activities, organizations and participants. Data management is just one piece of that complex system. Yet, as we have documented in this brief, data quality is essential to the integrity of a QRIS and ultimately to the effectiveness of the ECE system. This brief documents the challenges of data management in ECE data systems and provides concrete guidance for creating the infrastructure and governance that can address these challenges.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 12

Appendix A: Protocol for Interview with Selected States

[Introduction]

1.   NAME:

2. ORGANIZATION:

3. REVIEW QRIS DESCRIPTION FROM COMPENDIUM AND ASK FOR CHANGES SINCE EARLY 2010:

4. QRIS ROLE AND YEARS IN ROLE:

5. ROLE IN QRIS DATA MANAGEMENT:

[Review what we mean by data management]

1.   Who in state has most comprehensive knowledge of data management?

2.   If not respondent, get:

a.   Name

b.   Contact information

[State QRIS data system]

6. DO YOU HAVE WRITTEN PRINCIPLES, POLICIES, OR PROCEDURES FOR COLLECTION, ENTRY, AND MANAGEMENT OF QRIS DATA? Yes/No.

a.   If yes, would you be willing to share it with us? Please describe, including the successes and challenges.

b.   If no, do you rely on an already existing document (such as?) to guide you in data collection, entry, and management? Please describe, including the successes and challenges.

c.   If no, do you feel a need for such a document for QRIS?

7. PLEASE DESCRIBE EACH DATABASE IN WHICH QRIS DATA IS STORED.

Do you have a graphic or a visual image/flow chart depicting the different data bases used in the QRIS? If yes, will you be willing to share it?

Name of database	How the data are used in QRIS	Organization responsible for database	Key data contact at that organization (name and contact information)	Who makes decisions for this database?

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 13

7a. Please describe the extent to which unique identifiers have been created:

Level	Included in one or more databases	If included, extent to which the same identifier is used across databases	Comments
Facilities			
Workforce members			
Classrooms			
Children			

8.   WAS THE QRIS DATABASE SPECIALLY DESIGNED OR WAS IT BUILT USING AN EXISTING SOFTWARE PROGRAM?

a.   If built using an existing software program, please name it.

b.   How was the structure of the database determined?

c.   How are IDs given? Are IDs connected to other systems, such as the K-12 system and social services?

[Description of data collection, entry, and management]

9.   PLEASE DESCRIBE DATA COLLECTION AND ENTRY, INCLUDING THE SUCCESSES AND CHALLENGES.

a.   Are data collected over time? If so how is the date of data collection captured?

b.   How do you handle updates of the data (e.g., are previous data over-written)?

c.   How is QRIS data backed up and secured?

10.   WHAT PROCESSES/STEPS DO YOU USE IN MANAGING THE QUALITY OF THE DATA (E.G., RANDOM CHECKS, TESTING FOR ERRORS IN DATA, ETC.)? Please describe, including the successes and challenges.

a.   Verifying the data collected?

b.   Entering the data?

c.   Checking the data that are collected?

d.   It would be great to know more about the supervisory structure – is there someone who oversees the process, answers questions?

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 14

11. IS THERE A CODEBOOK FOR YOUR DATA SYSTEM? If so, how is this codebook managed and verified (e.g., who is responsible for the codebook, how often updated, electronic version)? Please describe, including the successes and challenges. Would you be willing to share the codebook?

12. HOW ARE INDIVIDUALS TRAINED IN QUALITY DATA MANAGEMENT (E.G., COLLECTING, VERIFYING, ENTERING, CHECKING, AND ANALYZING THE DATA, ETC.)? Please describe, including the successes and challenges.

a. Do individuals receive "refresh" training on their data management responsibilities? If so, who conducts these trainings?

HOW DO YOU GET FEEDBACK FROM INDIVIDUALS/UNITS/DEPARTMENTS USING THE DATA? Please describe, including the successes and challenges.

Appendix B: State and Local Area Examples

B.1 Ohio Step Up To Quality (SUTQ)

Ohio's Quality Rating and Improvement System, Step Up To Quality (SUTQ), uses multiple sources of data from several state systems. The SUTQ Database was built by Ohio specifically for the SUTQ and pulls data from the child care licensing, publicly-funded child care data systems. The SUTQ Database also stores all rating verification visit information. Ohio initially built a database for SUTQ, pulling data from child care licensing and CCIDS (Child Care Information Data System). In addition, their QRIS uses data from the professional development registry; however, the systems are not linked.

The biggest challenge Ohio faces with the QRIS database is that the foundation was built on an older platform of Microsoft Access. It requires on-going maintenance which includes re-builds in order to function properly. The agency standard is now a newer version of Microsoft and the database is not fully supported within the agency, and for that reason the QRIS is dependent upon third-party support.

The decision has been made to build a new web-based system as part of Ohio's Early Learning Challenge Grant in collaboration with the Ohio Department of Education (ODE). Through these efforts, Ohio will expand and re-engineer multiple non-connected data systems currently being used for licensing and SUTQ to create a modern, web-based system that links licensing data and that will be used by both Ohio Department of Job and Family Services (ODJFS, administrator of SUTQ) and ODE. The data system will include quality and regulatory data for all setting types, including small family child care homes. ODE and ODJFS will be able to upload licensing data and QRIS data for monitoring and reporting functionality. The new QRIS system will be built upon standards and criteria agreed to by ODE and ODJFS.

Ohio has convened a QRIS team with representation from both ODE and ODJFS to serve as the governance body for the new system. The QRIS team established a stakeholder group that included representation from the provider community and other early learning initiatives across the state. The Ohio QRIS team finalized the state quality standards and criteria that are being implemented through the new QRIS system, and they have set a vision and goals. They also selected data governance methodology and consolidated the QRIS system for Ohio. ODJFS established an enterprise data governance structure for the agency. The initial QRIS portion of the system is functioning. Additional functionality for SUTQ will be added every few months while the licensing portion of the data system is being designed and built – scheduled for launch in 2015.

For more information, see http://jfs.ohio.gov/cdc/stepUpQuality.stm.

B.2 Miami-Dade, Florida -- Quality Counts

Miami-Dade County's QRIS system is called Quality Counts and their data system is Miami WELS (Web-based Early Learning System), which is funded by The Children's Trust through a subcontract with the QRIS administrator, the Early Learning Coalition of Miami-Dade/Monroe. The data elements for Quality Counts are collected in a variety of ways through several data interfaces:

- Registry: individual education and professional development information, including practitioner demographics, career development goals and plans, scholarship usage, wage supplements received, and career advisors' time and activities

- Florida Department of Children and Families (DCF): child care licensing, including compliance and

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 16

violations and program size, location, and accreditation

■ Branagh ERS Assessment System: information about ERS scores

■ WELS: information about CLASS Pre-K and Toddler scores

■ WELS Provider Portal for program directors and family child care providers: to apply for Quality Counts, submit self-study information which includes ability to upload sources of evidence documentation and program quality improvement plans (with TA specialists), and apply for grants

WELS is the central database system for Miami-Dade's Quality Counts program. Memoranda of understanding (MOUs) are in place between WELS and the partnering agencies, Branagh Software System and the Children's Forum ("the Registry"). MOUs address ownership of the data, transmission procedures, frequency of data transmissions, costs if applicable, and security measures taken to interface data in a secure manner.

The Registry, ERS Assessment, and DCF system are interfaced based on unique IDs and license numbers given to providers; thus, errors can be captured. The subsidy data is interfaced twice a year, but other data is interfaced at every renewal point. WELS data views are current views, so the information is over written, but the history is backed-up and secured on a network in North Carolina.

The Registry provides a center report which generates the data for the staff qualification standard. This data is interfaced to WELS on a nightly basis. WELS pulls public data information from the DCF database on a nightly basis to maintain current information for all of Miami-Dade County providers. The Branagh Software System interfaces environmental rating assessments (scores and reports) on a nightly basis. All informational data is a one-way flow into WELS in order to maintain a centralized data system. WELS provides reporting capabilities for internal and community reporting.

Information from the Registry and Branagh ERS Assessment system is transmitted to the WELS database system to generate ratings. The ratings are reviewed and confirmed by the Rating and Support Specialist (RSS) team that processes ratings daily. The RSS team also manages the daily implementation and workflow of the data entering the QRIS system and ensures that entered or merged information is accurate through error reports that are generated. This team is also peripherally supported by content managers of each data system. Interfaced data are cross-checked with the data in WELS to determine accuracy of processes. If the data does not appear or there are errors, the project manager is alerted. There are also "exception reports" if an error occurs during transmission of any of the nightly interfaces.

For more information about this process, see http://welsfoundation.org/.

B.3 North Carolina License System

North Carolina's quality rating improvement system is licensing, which is directed by the Division of Child Development and Early Education (the Division). Data that is in this system is what is used to determine the star ratings for licenses that programs receive. Data related to education levels of staff, environment rating scale scores, child-staff ratios, and basic health and safety requirements are keyed, then the data system computes the level of stars earned. Data integrity is critical due to the linkages between NC's QRIS and licensing , and because the level of license is tied to subsidy reimbursement. Furthermore, data about

Best Practices in Data Governance and Management for Early Care and
Education: Supporting Effective Quality Rating and Improvement Systems

Page 17

programs' licenses and star ratings are on a public website used by parents and the community on a regular basis to make decisions about child care.

North Carolina created a licensing database in 1992 to maintain records related to visits, complaints, administrative actions, operators and licenses. Initially, this information was keyed at the state office based on paper documents that were created by statewide field staff. Subsequently, this system was revised to be a web-based system and a subset of fields were pulled to create a laptop system for field staff to use as they made visits to child care programs. The laptop system allows consultants to complete a visit report while in the child care program, print out the visit summary for the operator, and process an automated upload of data into the main system so data are immediately available.

The Division realized that in order to ensure data integrity, it was essential to have a training process in place as well as a quality assurance process. A training program was developed and a staff person ("Connectivity Consultant") was allocated to oversee this process, which is funded with the federal CCDF grants funds. In addition, a training site was developed that mirrored the production database but allowed for entry to be done and reviewed for accuracy without any actual impact on a program. The responsibilities of this staff person are to create and maintain the training documents and protocol, provide in-person training with staff, review data that was entered in the training site, ensure all types of training packets were completed, and verify that the specified level of accuracy was earned before the consultant was released in to the production system. On an ongoing basis, random packets are reviewed by clerical staff to ensure quality. The connectivity consultant remains available to field staff for questions, retraining, and any requested oversight. In addition, several training clips were developed that could be watched on demand, about procedures and questions that come up often or have high error rates. Data keyers and managers can review the clip at any point if they are uncertain about the steps they need to take for specific items.

When information about a program changes, the new information is keyed by program staff and new license generated, if applicable. All data are stored in a data warehouse that allows for generation of ad hoc reports for planning and evaluation purposes.

The value of this data integrity process is that it ensures that the Division, programs, families, communities, and researchers have confidence in data as it is used for policy development, law and rule change proposals, and research projects.

For more information about this process, go to: http://ncchildcare.dhhs.state.nc.us/general/home.asp.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 18

Appendix C: Glossary of Terms Related to Coordinated Data Systems

Central repository: A central repository contains data from more than one system. It allows data from multiple sources to be displayed or gathered in a single repository. It enables the retrieval of person- or organizational-level data that is stored in multiple databases.

Dataset: A set of data elements collected and shared to address particular needs; dataset usually refers to the content and may be a subset of a database

Database: Database refers to a set of integrated files.

Data models: Data model describes the file/table structures that will support the work you are trying to do in a shared environment, such that each instance of a data element is captured only once (called 'golden' and considered 'truth') from the source and then used wherever it is needed. A data model describes how the cross-sectional and longitudinal tables of data relate to each other to address the information needs of the organization. The data model needs to be built so that data can be captured and retrieved on individuals or organizations at a point in time (cross-sectional) or over time (longitudinal). Also, the file structure schema needs to be efficiently designed to capture both demographic and longitudinal data that uses metadata so that simplified files are possible. An efficient data model uses metadata to allow users to capture diverse kinds of information on an individual or organization in a simple manner.

Data sharing: Data sharing connotes more than interoperability does. Interoperability relies on standards to technically exchange data but does not deal with the permission to do so. Data sharing deals with those permissions in addition to making sure the data is in a form that makes sharing possible. One challenge to data sharing arises when permission to use data is only extended to the provider who collects the data. Data sharing requires double permission so providers can share further (e.g. a program creates a dual permission form that enables information to be shared with an additional program).

Data warehouse: A data warehouse is a centralized repository, but not all centralized repositories are data warehouses. A data warehouse is a physical location that stores data from multiple information sources and from which meaningful data for decision-making can be retrieved.

Federated database system: A federated, or shared, database system keeps component databases autonomous. It uses a defined architecture and interconnects databases so as to support partial sharing and coordination among multiple component database systems.

Governance: Governance is where stakeholders come together to make decisions about what the vocabulary will be, which nationally-recognized standard will be used for its representation, and who will have permission to access the data. It also governs the metadata (categories, such as education and training).

Information science: The science of capturing, managing, analyzing and using data

Informatics: Informatics adds domain knowledge to information science and hence is an applied science. Information science provides the theoretical underpinning that Informatics uses in practice. Practically all sciences are adopting informatics. It started in health, where there was a need to manage large amounts of data in order to serve patients, and now universities offer informatics degrees for a range of disciplines.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 19

Interoperability: Interoperability describes "the ability of two or more systems or components to exchange information and to use the information to make better decisions." (Johnson, 2011).

Metadata: A set of data that describes and gives information about other data

Point solution or point product: A term used to describe software that is designed to solve a particular problem without regard to related issues. Point solutions are widely used to fix a problem or implement a new service quickly.

Source databases: Within a coordinated data system, the database that is closest to the data is the source data system. Source or program databases are created and managed to meet program needs. Examples include licensing, registry, subsidy, and other early care and education databases.

Standards: Many types of standards are important in data management, including data vocabulary standards, data file structure standards, and governance standards. Two federal laws, the Family Educational Rights and Privacy Act (FERPA), revised 2011, and the Health Insurance Portability and Accountability Act of 1996 (HIPAA), tell states what they have to do but no standards are given—states have to observe the restrictions but how they do so is left to the states and to individual programs. Having clear standards is a critical piece of system-building.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 20

References

Burchinal, M. R., & Neebe, E. (2006). I. Data management: Recommended practices. *Monographs of the Society for Research in Child Development,* 71(3), 9-23. doi: 10.1111/j.1540-5834.2006.00402.x

Caronongan, P., Kirby, G., Malone, L., & Boller, K. (2011, August). *Defining and measuring quality: An in-depth study of five child care Quality Rating and Improvement Systems.* OPRE Report 2011-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Data Quality Campaign. http://www.dataqualitycampaign.org/blog/2012/12/unpacking-p%E2%80%9320w-data-governance/

Early Childhood Data Collaborative.(2010, August). *Building and using coordinated state early care and education data systems: A framework for state policymakers.* Retrieved on December 22, 2011 from http://www.dataqualitycampaign.org/files/DQC%20ECDC%20WhitePaper-Nov8.pdf

Federal Student Aid (2007, June 29). *Enterprise Data Management Data Governance Plan Version 1.0.* Retrieved on March 15, 2012 from http://studentaid.ed.gov/es/sites/default/files/fsawg/static/gw/docs/ciolibrary/ECONOPS_Docs/DataGovernancePlan.pdf

Friese, S., King, C., & Tout, K. (2013). *INQUIRE Data Toolkit.* OPRE Report # 2013-58. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. See http://inquiredatatoolkit.org/ for further information.

Friese, S. Tout, K., & Kirby, G. (2014). *Best Practices in Ensuring Data Quality in Quality Rating and Improvement Systems (QRIS).* OPRE Research Brief #2014-47. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Johnson, G (2011). *Implementing interoperability: The road ahead.* Presentation at the National Black Child Development Annual Conference, Nashville, TN.

Malone, L., Kirby, G., Caronongan, P., Tout, K., & Boller, K. (2011). *Measuring Quality Across Three Child Care Quality Rating and Improvement Systems: Findings from Secondary Analyses.* OPRE Report# 2011-30. Washington. DC: US. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Mosley, M. (Ed.) (2009). *DAMA-DMBOK functional framework*. DAMA International. Bradley Beach, NJ: Technics Publications, LLC.

National Center for Education Statistics (2014). Retrieved from: http://nces.ed.gov/programs/slds/resources.asp

O'Carroll, P. W. (2003). Information architecture. In O'Carroll, P.W., Yasnoff, W.A., Ward, M. E., Ripp, L.H., & Martin, E.L. (eds). *Public health informatics and information systems*. New York: Springer-Verlag.

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of Quality Rating Systems and Evaluations.* Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems

Page 21