OPRE Research Brief #2014-47
June 2014

Best Practices in Ensuring Data Quality in Quality Rating and Improvement Systems (QRIS)

Best Practices in Ensuring Data Quality in Quality Rating and Improvement Systems (QRIS)

OPRE Research Brief #2014-47
June 2014

Sarah Friese and Kathryn Tout, Child Trends
Gretchen Kirby, Mathematica Policy Research

Submitted to:
Ivelisse Martinez-Beck, PhD., Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP23320095631WC

Project Director: Kathryn Tout
Child Trends
7315 Wisconsin Avenue
Suite 1200W
Bethesda, MD 20814

This report is in the public domain. Permission to reproduce is not necessary.

Suggested citation: 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.

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).

Disclaimer: The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

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

Collecting and using data are core activities in a well-functioning Quality Rating and Improvement System (QRIS). Yet, data used in a QRIS are frequently housed in different systems, using different data management techniques. Ensuring a high level of QRIS data quality involves implementing a number of best practices drawn from established practices used in other fields.

The purpose of this brief is to describe the specific strategies QRIS data stakeholders can use to improve upon the collection, management, and dissemination of QRIS data. The audience for this brief includes QRIS program administrators, technical assistance providers, data managers, and researchers.

This brief is structured around the five stages of the Data Lifecycle: planning, collection, processing, management and distribution. Best practices are recommended for each stage of the Lifecycle.

Planning Best Practices

- Decide which data elements to collect.
- Plan for data to examine QRIS validity.
- Identify existing systems with which to link.
- Establish a system of Unique Identifier Codes (UICs).
- Develop an infrastructure for training, supervision, and technical support for staff with responsibility for data collection and data management.

Collection Best Practices

- Use web-based, on-site data collection methods.
- Collect data about all programs and practitioners.

Processing Best Practices

- Input data at the rawest level possible.
- Minimize overwriting of historical data.

Management Best Practices

- Maintain up-to-date codebooks.
- Systematize management so that knowledge is institutional rather than person-centric.

Distribution Best Practices

- Ensure that data releases adhere to privacy and confidentiality standards.
- Create an automatic feed of ratings data to avoid errors.

The best practices recommended in this brief reflect some of the most salient issues in QRIS data quality but are not inclusive of all data quality practices. States and localities implementing the data quality activities described in this brief should work closely with technical architects and data managers, whether internal or external, to determine which best practices would be most useful for QRIS implementation.

Best Practices in Ensuring Data Quality in Quality Rating and Improvement Systems (QRIS)

Introduction

Data on early care and education systems are typically collected, housed and maintained by different state and local agencies at different levels of the system (individuals, classrooms, programs) for multiple purposes. In recent years, Quality Rating and Improvement Systems (QRIS) for early care and education have become key producers and consumers of data. In a QRIS, data are used to assess the quality of programs serving young children, to assign and disseminate quality ratings, to support the provision of technical assistance to improve program quality, and to facilitate monitoring and evaluation of QRIS activities. QRIS data are visible publicly—through an assigned rating of quality—making the use of data distinct from other quality improvement or professional development initiatives.

The centrality of data in a QRIS elevates promotion of data quality as a primary task for QRIS administrators, contractors and evaluators. Yet, despite the wealth of information that exists about the importance and implementation of rigorous data management practices for business and governmental organizations, specific guidance to support application of these practices to QRIS is lacking. Established data quality and management practices can be applied to the collection and use of data for the unique purposes of a QRIS to improve the quality of the data in the system. This brief aims to support QRIS data quality by providing an overview of best practices in data management for QRIS at each phase of the data lifecycle.

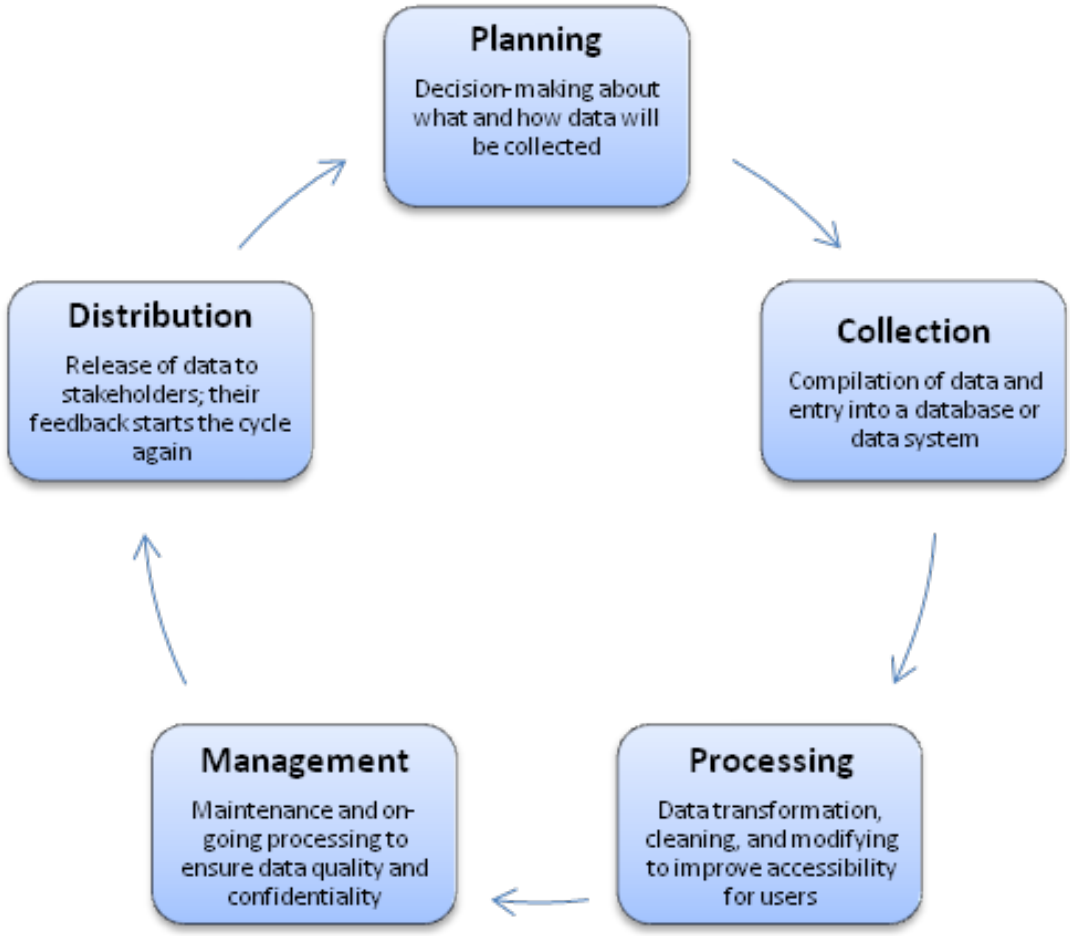
This brief describes the specific strategies QRIS data stakeholders can use to improve upon the collection, management, and dissemination of QRIS data; it does not focus on the development of a data system infrastructure. The primary audiences for this brief are QRIS program administrators, technical assistance providers, data managers and researchers.

A companion brief covering the topics of data systems and governance, Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems (Weber & Iruka, 2014), can be used in conjunction with this brief for a fuller picture of the data landscape for QRIS. Readers can also consult the *INQUIRE Data Toolkit* (2013) and other resources developed with the support of the Office of Planning, Research and Evaluation (OPRE) in the Administration for Children and Families in the U.S. Department of Health and Human Service for further guidance on QRIS data issues. For more information about Early Childhood Integrated Data Systems (ECIDS), please see resources developed by the National Center for Education Statistics (2014).

Data Lifecycle

The data lifecycle provides a framework (see Figure 1) to address data quality practices in QRIS for each stage of the lifecycle (Data Documentation Initiative, 2009). The data lifecycle begins with planning, progresses through the subsequent stages of collection, processing, and management, and ends with distribution. The data lifecycle is not finite; experiences at each stage of the lifecycle ultimately inform the next planning stage to begin the cycle anew. Best practices for each stage can be tailored to provide specific guidance for QRIS data.

Figure 1. The five stages of the data lifecycle



Source: Adapted from the Data Documentation Initiative

The data lifecycle can be applied to a QRIS at all stages of implementation, including in development, in a pilot phase, or in mature operation. While the term "lifecycle" implies a systematic progression through the stages (one completes before the next one starts), we recognize that in mature data systems that have already experienced numerous cycles, the stages of the lifecycle can occur concurrently, both informing and being shaped by the knowledge learned during other phases.

In the sections that follow, we discuss QRIS data management practices for each stage of the lifecycle. Discussion of each stage begins with a goal of implementing the best practices and then highlights select best practices within a QRIS. This brief does not provide an exhaustive list of high-quality data practices. Additionally, many of the best practices described here are not the exclusive concern of the portion of the lifecycle in which they are presented; rather, many of these best practices are interrelated and applicable at more than one stage. Planning for specific QRIS issues may benefit from the support of experts who can provide a full set of recommendations that are appropriate for the unique needs of each QRIS in the context of the broader system.

Best Practices for Ensuring QRIS Data Quality at Each Phase of the Data Lifecycle

Planning

Goal: Ensures the universe of data collected meets program management and research needs, acknowledges the need to link data sets, and establishes flexibility for responding to unexpected data challenges

Critical to the planning stage is to identify the data needed to address the core functions of a QRIS—assessing quality, assigning ratings, disseminating information about quality, informing quality improvement, monitoring progress and evaluating effectiveness. During the planning phase, both programmatic and technical data stakeholders engage in a process to map out what QRIS data to collect, how to collect it (the source of each data element as well as where it will all be housed), and how to connect the QRIS data system with other important data (e.g. child care licensing, workforce registry) that may be used to determine program ratings.

Best practice: Decide which data elements to collect.

The primary function of QRIS data is to produce a rating for early care and education programs that conveys information about program quality to parents (as consumers) and that can be used to target quality improvement efforts. At a minimum, the data elements to collect are the indicators that feed into the calculation of quality rating levels.

Available resources that draw from the education field and the experiences of existing QRIS can provide guidance on a starting set of core data elements to include in a QRIS and in the early care and education system overall. Adhering to data standards such as those available from the *INQUIRE Data Toolkit* (2013) and the *Common Education Data Standards* (2014) can ensure collection of the appropriate elements, at the right level (e.g., child, family, practitioner, program site, organization, system), using the most common variable options. Common data standards for early learning are not intended to be all-inclusive or prescriptive; each state or locality will need to add to or modify the data elements to meet their distinct needs.

Use of a logic model can assist in the application of the general guidance available to meet the needs of a specific QRIS. A logic model identifies the essential activities and expected outcomes of the QRIS. Guidance for logic model development as well as examples are available in the QRIS Evaluation Toolkit (Lugo-Gil et. al, 2011). Planning at the outset of QRIS development to collect data elements that align with the activities and outcomes will save resources and time at later points in the data lifecycle. Completing this process helps to identify not only which elements to collect but also the level at which to collect them. For example, data to produce quality ratings may be needed about individual staff (such as education and training), the classrooms in which they work (such as child-staff ratios or measures of classroom quality), and the programs housing the classrooms (such as management policies and practices or methods of family engagement). Planning efforts can also identify ways to reduce duplication of data and determine strategies for linking data across different levels.

Best practice: Plan for data to examine QRIS validity.

As QRIS grow and expand with the use of public dollars, it is increasingly important to validate that the systems are doing what they intend to do—measuring and improving quality in early care and education programs. As relatively new systems, definitive answers about the "right" indicators and the "right" measurement methods are still lacking. Given the nuances to the context and goals of each QRIS, there is not a one-size-fits-all approach. Each QRIS needs to have the data to assess its own validity and, in turn, can contribute to the knowledge base for the future.

Validation is an essential activity supporting continuous quality improvement of QRIS because the results of validation analyses can be used to revise or restructure QRIS quality indicators and rating processes. "Validation of a QRIS is a multi-step process that assesses the degree to which design decisions about program quality standards and measurement strategies are resulting in accurate and meaningful ratings" (Zellman & Fiene, 2012, p.1).

Comprehensive data for validation can include the full range of indicators within the rating structure and should be collected for all participating programs. Often, data are not collected systematically across all programs because certain elements may not be needed for some programs to produce a rating level. However, the lack of consistent data across programs inhibits validation analyses without new data collection. For example, QRIS that use a block structure to assign ratings requires that a program meet a full set of indicators at a given quality level before they can be rated at that level. When it is clear that a program is not going to meet all the criteria at higher levels of the QRIS, data collection typically stops for that program. This is particularly true for systems that allow programs to set rating goals: programs select their goal and only submit the documentation needed to meet that goal. A lack of data about whether the program meets quality criteria at a level above the designated rating will constrain data analyses to determine if the current configuration of indicators and quality levels is working as intended. Whether or not program ratings should involve "scoring up" is important to consider, especially early in the implementation of new quality criteria.

Another practice that can support validation analyses is to collect and store data at its rawest—or most basic—level. For example, data from observational tools will be most useful at the level of individual indicators and items, rather than at the level of a summary score. Storing data at this level allows analysts access to more information about site-level variation and patterns of scoring on particular domains of quality that may provide insights into how well the QRIS is working to differentiate quality among early care and education programs.

Planning for the collection of actual measures of indicators rather than summary information (check boxes or scores) is vital to supporting validation analyses. For example, in a block system, raters could just check "yes" or "no," indicating whether a program meets each indicator. These data may serve the program purpose to produce an overall rating but will not always provide the information needed for validation. Validation relies on the measures or scores for each indicator for each participating program. When assessing teacher education level, a "yes" can indicate that the lead teacher has, for example, at least a CDA. But, it does not provide information about the level of education the teacher may have beyond a CDA (that could meet the next-highest rating level criteria for this indicator). Similarly, in a point system, assigned points can be indicative of the level of education, but still may be inexact. Collecting the specific information for each indicator and then determining a yes/no or points will serve both the program and evaluation functions for the data.

Best practice: Identify existing data systems with which to link.

Some data on indicators to produce a quality rating may already be collected in different systems; work in the planning stage is needed to determine how to make the best use of existing data, and how to do so efficiently. For example, data on programs' licensing status is frequently housed in a system that is separate from the planned (or existing) QRIS data system. QRIS stakeholders can plan ahead to determine methods of accessing information about licensing status, or other existing systems. A paper transfer of the data needed for rating is not an ideal option for efficiency or possibly for maintaining data security, particularly for data on individuals. The most efficient long-term strategy is to develop electronic data-sharing methods whereby systems "talk" to each other to exchange information.

Best practice: Establish a system of Unique Identifier Codes.

The best practice for identifying the programs, practitioners, and children with data in a QRIS is to have unique statewide identifiers that are used across systems. For example, participating programs would have one program ID that would be used across the QRIS, licensing, and child care subsidy data systems. Practitioners in a registry would have their own unique ID that would link them to their program through the program's ID. Similarly, a unique identifier for children can facilitate the transfer of information about their experiences across systems (e.g., QRIS, subsidy, education, health) to improve planning and service delivery and, ultimately, to track them through different programs over time (for example, as they move through different QRIS programs and into kindergarten).

Establishing a system of unique identifiers that are used across data systems requires long-term planning that many states find difficult to implement even with the appropriate resources in place. In lieu of being able to implement system-wide unique identifiers, it is important to develop a plan to link the identifiers already in place or to create a system where identifiers build on existing identifiers such that linkage is made easier. For example, a licensing ID may be included in the QRIS database with the program's QRIS ID. For children, matching software may be used to link information from the QRIS to other data systems using different identifying information.

Best practice: Develop an infrastructure for training, supervision, and technical support for staff with responsibility for data collection and data management.

Quality assurance of data used in a QRIS relies on implementation of a strong infrastructure for the training and supervision of staff. Protocols to promote consistent and comprehensive data collection provide a necessary foundation, but staff training in use of the protocols is critical to data quality. Pre-service

orientation and training can focus on both specific skills, such as the use of technology for data collection in the field, and on instilling a substantive grasp of the purpose and content of data collection more broadly. It is particularly important to provide training and support for staff who may be acclimated to different data collection protocols or different standards of data quality they used on projects prior to their role in QRIS so that they don't transfer those other practices over to new data collection efforts. In planning training, important elements that promote consistency and reliability for data collection and entry include descriptions of data entry guidelines, support on data collection technology, expectations for timing of data entry, and rules for data checking. Easy access for staff to written documentation can further promote adherence to data collection protocols, and avoid having staff make off the cuff judgments because quick information is not available. Monitoring protocols for the regular review of data quality by supervisors or managers is also important to ensuring quality across all the stages of data collection as well as across the variety of staff who enter data.

Collection

Goal: Maximizes efficiency, promotes interest in linking across programs, and minimizes the opportunity for error

In the collection phase, information used to make ratings determinations is captured and entered into a database or data system using a variety of methods that range from simple to sophisticated. Simple data collection procedures require program staff to complete paper forms that are sent to the agency or organization responsible for the QRIS. Agency staff then hand-enter the data into an electronic spreadsheet or web-based data entry system. Simple data collection procedures are time- and resource-intensive and more prone to error than sophisticated systems. In sophisticated data collection systems, programs enter their ratings data directly into a web-based system that is programmed to automatically calculate ratings. Scoring of observational assessments is also automated. Sophisticated systems still rely to some extent on non-automated processes, because staff will need to verify the data entered into the system for some percentage of programs. However, sophisticated systems, while they have greater development costs upfront, are often more cost-effective and less prone to error in the long run.

Best practice: Use web-based, on-site data collection methods.

Past standard practice in QRIS data collection relied on paper collection with electronic data entry at a later date. Technological advances make the better practice of direct data entry into web-based systems much more feasible. On-site, web-based data collection reduces the incidence of data error by eliminating the transfer from paper to an electronic format. An additional benefit to this type of data collection is that it decreases the time burden on the QRIS by merging the two separate steps of data collection and data entry into one.

Even with the benefits web-based systems can provide, there are important considerations that come with the use of this format. First, clear training and guidance for look-up practices are needed to ensure that the data are connected to the correct child, classroom or program. The risk of incorrectly assigning scores is greatest for organizations that are managing multiple sites (including licensed child care centers, Head Start programs, and school-based programs), where there may be confusion surrounding which classrooms are associated with each site. Second, attention to data privacy and confidentiality concerns are crucial when using web-based data collection methods. Web-based data entry applications should require authentication credentials that vary based on the type of user (e.g., different levels of access for program staff and QRIS staff) and adhere to rigorous data security standards. Data entry forms should be created such that staff or parents at the program being observed cannot view final calculations of scores.

Best practice: Collect data about all programs and practitioners.

Many QRIS data systems capture information about only a portion of the total possible population. For example, some workforce registries are voluntary and contain information about only those who have completed the necessary steps to enroll. When a QRIS requires that practitioners enroll in the registry, participation may increase. However, these registries still do not contain data on the full universe of practitioners. Additionally, registries are often missing practitioners that have previously been required to maintain learning and education records in another system like K-12 data systems.

To improve coverage of workforce registries and the ability to use the data to answer questions about the workforce at a state level, it is important to prioritize consistent participation in registries. Collecting workforce data universally decreases the burden on programs to have all of their staff enroll in the registry in a short amount of time as they begin their participation in the QRIS. States also benefit from having professional development data on a wide variety of practitioners to help them understand the differences between those working in rated and unrated programs. Finally, registries can be a useful personal tool for all practitioners to track their education and training as their careers progress.

Processing

Goal: Ensures consistent use of the data and identification of data errors

During the processing phase, data that have been collected are cleaned and transformed to make the data usable to agency staff operating the QRIS, researchers, technical assistance providers and policymakers. It is in this stage that raw data about program quality (for example, the individual items on observation measures or achievement of specific indicators) are processed and converted into a final rating score. Because quality rating scores are available publicly and intended to inform parent decision-making and guidance on quality improvement, careful attention needs to be paid at this stage to ensure that the final quality score is an accurate reflection of a program's unique combination of sub-scores and indicators achievement.

Best practice: Input data at the rawest level possible.

As noted in the planning stage, QRIS data are most useful and accurate when entered into the data system at the rawest, most disaggregated level possible. The rawest level means the level of actual measurement, whether observed or documented (such as teacher education level or child-staff ratios) rather than a summary score (such as a scale for observational quality, a yes/no for meeting an indicator, or assigning points). In addition, data collection at the rawest level improves data quality because it limits interpretation or error in producing a summary score and promotes reliability in the creation of summary scales, scores, and rating levels. Collection of raw data supports automated processes for scoring indicators—whether in a block, point, or hybrid system. And, raw data collection eases transitions when QRIS change scoring procedures, cut points, or even switch between block and point systems. The raw data can be manipulated and calculated as necessary to accommodate any system change, and can best support validation and research purposes.

Data that is collected at the item level is often entered into a database at the subscale or scale level in the interest of saving resources or because it is decided that item-level data is not needed. This is a common data entry practice for observational quality data. The drawback of not entering and storing data at its rawest level is that it becomes difficult to understand variations in scoring that are obscured by using a

summary score. Researchers charged with QRIS validation are limited in analysis options if they are unable to access data at the indicator level.

One way to efficiently ensure access to data at the rawest level is to use on-site web-based data collection and software that automatically calculates higher-level scores. For observational measures of quality, this means the observer enters items directly into a system that is programmed to calculate subscale and total scores. With this type of system, the data should be accessible at the raw level when needed. Proprietary systems can present a challenge to this if they limit the way in which users are able to access the data they have entered. During the planning phase, attention must be paid to ensuring that data ownership remains with the QRIS-responsible agency and not with the owner of the proprietary data entry system for observational quality.

Best practice: Minimize overwriting of historical data.

During the processing stage, internal decisions that were made about what data to collect and how to collect it are implemented. New fields must be created to house data that have not been collected previously. For example, if programs go through the QRIS rating process every two years, all of the fields that were created for the first rating need to be replicated for subsequent ratings so that analyses examining change over time can be conducted.

Technical architects and QRIS program staff must work in concert to determine how the historical data will be housed. One option is to house it in the same data set. A benefit of this strategy is that the data remains in one place, but a drawback is that it creates a data set that can become difficult to manage. Another option is to partition the data set such that only the current data are housed in the main file and any historical data are kept in a separate file. Working with the main file becomes easier in this scenario, but additional work will be needed to merge the data sets to do historical comparisons.

Data lifecycle QRIS example:

Observational quality data

Planning - Observational data presents a challenge for management because it is typically collected on a classroom level and then combined across classrooms and "rolled-up" to the program level. To address this issue in the planning stage, it is helpful to create a system of unique IDs that allows for a direct connection between the classroom and program levels. One option would be to have separate types of IDs for classrooms and programs and to ensure that both IDs are housed in any data set with information about either. Another option is to embed the program ID into the classroom ID such that program membership is always clear. At this stage, it is also important to identify training procedures and monitoring processes to ensure that contractors implementing a portion of the rating system (such as the collection of observational data) are using rigorous data collection practices.

Collection - A best practice in data collection of observational quality data is to develop or use an existing system that allows observers to collect data directly into an electronic format during the observation. This process eliminates the need to transfer that data from a paper form and reduces the likelihood of error. An up-front investment will need to be made to develop a data entry system—whether one is purchased off-the-shelf or built in-house—that can house observation data prior to the start of data collection. This investment may seem substantial, but it dramatically reduces the amount of time and staff hours needed to collect, clean, and enter data collected on paper.

Processing - Entering observational data at the item level provides more fine-grained data that can be used for QRIS validation analyses. Item-level data can be used to examine patterns in scoring or correlations between scores on

Special attention needs to be paid to how proprietary QRIS data systems handle historical data. As a QRIS matures through many rounds of rating and re-rating, researchers and program staff will need access to the full history of programs' participation and change over time, not just the current rating cycle, in a format that allows for data analysis. Proprietary data systems will often include customizable reports that provide a quick snapshot of some component of the QRIS data. Reports are a quick and valuable tool for program administration, but they typically are not designed to include information that allows for more sophisticated data analyses. In addition to reporting features, researchers and agency staff need access to flat files with complete information to fully understand the impact of the QRIS over time.

Management

Goal: Promotes accountability and usability of data by multiple users

Data management is the process of tracking and storing data for long-term usage. Data management supports the ongoing understanding, use, and analysis of data by providing a map to unmanipulated data as well as standard created variables. As new rounds of data are collected, they are added to the existing data set. Creating thorough documentation of the data elements, a history of changes to the data elements and guidance on how to combine and use data elements is a critical activity at this stage. Thorough documentation allows the manager and users of the data to understand why data were collected and stored in a certain way and the justification for the creation of new variables.

Best practice: Maintain up-to-date codebooks.

Codebooks document the specifics of individual data elements and provide data users with details about changes that have occurred in data collection over time, to help them understand the characteristics of the data and how they can best be used. At a minimum, the codebook should include details such as a definition, variable type,

Data lifecycle QRIS example (cont.):

particular sub-scales and other quality indicators that would not be possible if only total scores were available.

Management - Codebooks detailing observational quality data elements should be clear as to the periodicity of the observations (e.g., once, every year during a specific time frame, every two years on a rolling schedule) so that users are clear about the relationship between the observational data, other quality indicators and overall ratings. Additionally, information in the codebook should provide basic details about any observational items from standard tools that are not scored and why new variables were created if they differ from the standard way of reporting sub-scale or total scores.

Distribution - Feedback reports are sometimes provided to programs following the completion of their observation. Distribution of feedback reports should be carefully monitored by program ID to ensure that no errors are made in distributing feedback to the wrong program. The feedback report should also be limited in scope, highlighting areas of strength and room for improvement rather than providing a detailing of the full score.

and variable options for each data element. Ideally, the codebook will provide data users with a clear sense of the evolution of a variable and its connection to other variables. This is especially important for data elements where the concept of the element remains the same but the way in which it is collected changes. For example, if a state QRIS has changed its structure for developing ratings (e.g., from points to block, points to hybrid, block to hybrid, etc.), the old rating structure data elements will exist in the dataset alongside the new data elements reflecting the new structure. The codebook should indicate the connection between these elements and the reason why the previous set of elements stopped being used in favor of the new ones.

Best practice: Systematize management so that knowledge is institutional rather than person-centric.

In the absence of a robust system of documentation including codebooks, knowledge about a complex data system become specific to the person whose job it is to manage the data. The danger of this arrangement is that when that person moves out of their role, their knowledge is lost and, even with a lengthy training period with a new data manager, it cannot be recovered by someone new. Details about data management should not rest with one person, even if day-to-day management of the system is the responsibility of one person. Robust documentation ensures that knowledge is institutionalized rather than person-centric.

One of the benefits of proprietary QRIS data systems is that the data system developer has a direct financial incentive to ensure that knowledge is not localized within one member of their staff, because the long-term sustainability of the system would be compromised. One drawback of proprietary systems is the consideration of what to do long-term if that business ceases operation. Even in circumstances where the company responsible for the data system is responsible for documentation, the QRIS-administering agency should maintain copies of that documentation should they need to refer to it at some point in the future and it is no longer available from the original source. Additionally, states using QRIS proprietary data systems will need to keep robust documentation detailing how that system is linked to other systems like a registry, child care licensing, or datasets with QRIS research and evaluation data.

Distribution

Goal: Disseminates information to data stakeholders using consistent processes that ensure privacy and confidentiality

During the final stage of the lifecycle, the data that have been collected, cleaned, transformed, and analyzed are released for use by data stakeholders including agency staff, researchers, policymakers, and parents. Access to the data will vary between the stakeholders depending on the information they need and the level of identifying information that the data users are allowed to access. QRIS agency staff may have access to all the data, as may staff in other governing agencies who need the data to fulfill statutory responsibilities. On the other end of the spectrum, parents and policymakers receive only processed and aggregated data that communicate quality levels at the program level, or that communicate system progress and goals. Researchers will often receive data that is somewhere between completely identified and de-identified, but that is not representative of the entire universe of information that was collected.

Best practice: Ensure that data releases adhere to privacy and confidentiality standards.

Identifiable QRIS data should be available to only a small, select group of users who need full access to ensure the program is operating successfully. Typically, data users with full access will be limited to a group

of staff at the agency or organization responsible for managing the QRIS, and possibly including researchers who are contractually assigned to evaluate the QRIS.

Any person or group with full access to the data must be able to ensure a high level of data confidentiality. In the context of QRIS, data confidentiality would include having a limited number of secure data files that connect child, practitioner, and program through their IDs, and strict adherence to data-sharing agreements between agencies that have access to full data sets. Special consideration needs to be paid to ensure that data that linkages across multiple levels (e.g., child, practitioner, program) and data systems (e.g., QRIS, registry, licensing) are compiled in a way that defaults to the strictest set of guidelines for any of the individual pieces of data. For example, if licensing data that are public are combined with child assessment data that are highly private, the rules governing the child assessment data should be the default assumption for the new data set.

Best practice: Create an automatic feed of ratings data to avoid errors.

Final ratings of program quality are available to the public through a website in most states. Early care and education programs recognize the high stakes nature of the ratings since they have a direct impact on the community's perception of program quality. Given the sensitive nature of ratings data, it is vital that any QRIS data available to the public are completely error free. Care must be taken at the processing and management stages of the data lifecycle to ensure those programs' final ratings are being made with accurate data.

To ensure use of clean, accurate data, ratings designated in a QRIS data system should be automatically fed to the public website displaying those ratings. Automatic feeds minimize the possibility of data entry error and ensure that data need only be checked in one system instead of two.

Conclusion

As states and localities continue to develop QRIS or engage in redesign activities, there is a need for guidance on strategies to assure QRIS data quality. This brief uses the data lifecycle as a framework for outlining essential activities and select best practices that facilitate data quality in a QRIS. As noted, the intention is to provide a starting point for data quality activities. Each state or locality embarking on development of a QRIS data system or engaging in revisions to existing data collection and management protocols will potentially benefit from the assistance of a technical advisory board and a consultant who can assess current practices. Nevertheless, the framework and practices described in this brief provide a useful reference and resource to support data quality assurance efforts.

References

Common Education Data Standards. (2014). Retrieved from https://ceds.ed.gov/Default.aspx

Data Documentation Initiative. (2009). *DDI Lifecycle*. Retrieved from http://www.ddialliance.org/

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.

Lugo-Gil, J., Sattar, S., Ross, C., Boller, K., Tout, K, & Kirby, G. (2011). *The Quality Rating and Improvement System (QRIS) Evaluation Toolkit.* OPRE Report #2011-31. Washington, DC: U.S. Department of Health and Hunan Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

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

Weber, R. & Iruka, I. (2014). *Best Practices in Data Governance and Management for Early Care and Education: Supporting Effective Quality Rating and Improvement Systems*. OPRE Research Brief #2014-35. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zellman, G. L. & Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for Early Care and Education and School-Age Care, Research-to-Policy, Research-to-Practice Brief.* OPRE Research Brief #2012-29. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.