The Optimal Reference Guide:

The Data Quality Manual

Data Quality Series – Part II

Extraordinary insight™ into today’s education information topics

By Glynn D. Ligon
Table of Contents

Introduction ............................................................................................................3
Steps for Achieving Data Quality .........................................................................3
Data Quality 101 .....................................................................................................4
The “don’ts” to avoid messing up your data ..........................................................4
Software Vendors ...................................................................................................7
The 80/20 Rule of Data Quality ..............................................................................8
Checklist for Sensing the Quality of Data .............................................................9
The Data Quality Rating Scale ..............................................................................12
The Four Great Truths about Data Quality ............................................................16
Steps for Ensuring Data Quality ..........................................................................16
A Final Note about Error .......................................................................................16
Conclusion ............................................................................................................16
Introduction

Let’s get to work and improve data quality.

*Data quality matters now.*

*Data quality is an official buzz word.*

*Data quality steps are known now.*

*Data quality is for everyone.*

*Data quality is detectable.*

*Data quality saves money.*

*Data quality relieves stress.*

By assimilating the conventional wisdom about data quality with the real school experiences of our ESP professionals, we have been able to create a tutorial on the practices that cause bad data and the processes that ensure quality data.

**Steps for Achieving Data Quality**
The authors assisted the U.S. Department of Education in the development of a set of data quality standards for program data. A training package was developed from those standards and sessions were conducted with program office staff. We took those relatively high-level standards and created a step-by-step process for managing the quality of data across an entire annual cycle.
Data Quality 101

The “don’ts” to avoid messing up your data

Never, ever create a reporting format that allows for:

- leading or trailing zeroes
- repeated numbers or letters in an identifier or code
- mixing numbers and letters in an identifier or code unless 0, O, I, 1, 1, and all other confused characters are left unused

The most frequent and insidious errors that plague an information system:

DO NOT:

1. Make notes in data fields.

First Name Field:
“Mandy (but mother says she prefers to be called “Pookey”)”

2. Copy and paste from one file (format) to another.

<table>
<thead>
<tr>
<th>Pat</th>
<th>M</th>
<th>Johnson</th>
<th>Jr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Johnson, Pat M, Jr

3. Be lackadaisical when the requirements are precise.

<table>
<thead>
<tr>
<th>Patrick</th>
<th>M.</th>
<th>Johnson</th>
<th>Jr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat</td>
<td></td>
<td>Johnson</td>
<td></td>
</tr>
</tbody>
</table>

4. Add codes to be more specific.

1 = Graduate
2 = Transfer
3 = Retainee
U = Unknown
M = Sent to Marie for Coding

5. Make the data your own.

Phone Number Field:
“555-555-5678 except on Tue then 656-555-5555”

6. Give everyone the same value just to fill the field.

SSN Field:
“111-11-1111”

7. Submit split or duplicate records.

<table>
<thead>
<tr>
<th>Student Name</th>
<th>Birthday</th>
<th>Test Score</th>
<th>Course Grade</th>
<th>Absences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat Johnson</td>
<td>09111999</td>
<td>98</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>Pat Johnson</td>
<td>09111999</td>
<td>98</td>
<td>A</td>
<td>3</td>
</tr>
<tr>
<td>Kelly Smith</td>
<td>12251999</td>
<td>79</td>
<td>B</td>
<td>8</td>
</tr>
<tr>
<td>Kelly Smith</td>
<td>12251999</td>
<td></td>
<td></td>
<td>8</td>
</tr>
</tbody>
</table>

8. Ask for forgiveness rather than permission.

“Oh, hello, yeah, I think I may have accidentally left all the Title 1 codes off my file. I’m really sorry. Can you ever forgive me?”

9. Argue with official names, spelling, or capitalization.

District Name Field:
“Colorado Springs”
(Official Name: El Paso County District 11)

10. Be right when the world is wrong.

Street Name Field:
“Arroyo Seco”
Arroyo Seca is the official name.

11. No matter how dumb they act, don’t say students were born yesterday.

Birth Date Field:
“April 11, 2008”

12. Be creative to get double use from the data.

Course Field:
“Lunch A”

13. Be better when the software is good enough.

Gender Field:
“Female”
(Valid Code = F)

14. Keep doing things the way you did before the new software was installed.

“My Excel spreadsheet is really the official record for my students.”

15. Call a friend at the district office or SEA and ask for her/him to correct your data.

“Hi Coleen, would you be a dear and just change those LEP codes for me again this year?”

16. Copy and paste without being extra careful.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Gender</th>
<th>First Name</th>
<th>Last Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>M</td>
<td>Freddy</td>
<td>Hanson</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>Sandra</td>
<td>Hernandez</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>Charlotte</td>
<td>Webster</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>John</td>
<td>Johnson</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>Michelle</td>
<td>Michelle</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>Juan</td>
<td>Paredes</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>Janelle</td>
<td>Smith</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>Herbert</td>
<td>White</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>Snoop</td>
<td>Perro</td>
</tr>
</tbody>
</table>

17. Think of data quality as an as-of-date requirement (wait to get everything right on the reporting date).

18. Pass data entry on to someone who doesn’t know the rules or can’t follow them.

“Our student aide will enter everything. Huh? FERPA? Training?”
Software Vendors

Software vendors are your partners. Better yet, software vendors are your “employees.” They need to make you look good. You must insist they follow the rules. Of course, this means that the people paying the vendors must insist. That may be the SEA or the LEA, or at times an individual school or program.

When we began the first statewide data collections using SIF in Wyoming, the “SIF certified” agents for student information systems (SIS) sent data to State Report Manager (SRM), ESP’s product for collecting and verifying data for the Wyoming Department of Education (WDE), using whatever codes they found in each district’s SIS. SRM’s business rules flagged them as fatal errors. This began a nationwide effort to accomplish two objectives. First, SIFA had to enhance their certification process to require that agents follow the complete standard including use of approved codes. Second, the SIS vendors had to enhance their agents to crosswalk or accept only approved codes. If the line had not been drawn in the sand at that point, the WDE staff would have continued to fix each district’s submission file before certifying the collection to be complete and ready for use.
The 80/20 Rule of Data Quality

You can either put in 80% of the effort cleaning up the data every year—or only 20% of the effort up front to establish clear rules and insist they be followed. Yes, that 20% is a lot of effort up front. Standard operating procedure is that work is done just good enough at each step because someone later on will clean things up if it’s really that important. That’s unacceptable. The 80/20 rule has been changed in Wyoming and other states using SRM as a gatekeeper for data quality to the 20/2 rule. That’s 20% of the effort is invested up front to ensure all business rules are met and only 2% of the effort from then on to handle outliers.

The greatest benefit has accrued to the local schools and districts. Using the specific, user-friendly edit reports that SRM provides as their trial data are tested, they have improved their processes to avoid entering or perpetuating many of the data problems that were inherent in the legacy systems. School and district people have been happy to improve once they received clear direction on where to change.

Process Flow of Reported Data:
- Declaring by the original source of the data (parent)
- Entering by the collector
- Compiling for reporting
- Sending
- Receiving
- Mapping
- Importing
- Accessing
- Analyzing
- Formatting
- Labeling
- Explaining
- Interpreting
- Using
Checklist for Sensing the Quality of Data

Sometimes the best way to determine the likelihood of quality data is for a human being to stare at the numbers and see if they make sense. Read *Blink: The Power of Thinking Without Thinking*, 2007, Malcolm Gladwell, to see how much of an expert you probably are when it comes to your own statistics.

From decades of proofing data reports, Figure 1 summarizes some ideas for checking the data for possible errors. Steps 1 through 12 are somewhat in order of their sophistication, but number 13 sums up the lesson from *Blink*—What’s your gut reaction?

**Figure 1  Steps for Validating Data**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Your Best Guess</td>
<td>Write down your best guess of what the statistic should be. How close to your prediction is the reported statistic? From all you’ve read, you know that reported dropout rates range considerably, but you expect the local rate to be about 3% a year. The preliminary rate sent to you from MIS is .35%. (Correcting an errant decimal made the rate 3.5%. That’s reasonable.)</td>
</tr>
<tr>
<td>2.</td>
<td>Prior Statistic</td>
<td>Find a previously reported statistic, preferably several across reporting times. How close to prior trends is the reported statistic? The prior four years’ dropout rates have been 6.7%, 5.4%, 3.8%, and 3.4%. So, 3.5% looks reasonable.</td>
</tr>
<tr>
<td>3.</td>
<td>Another Entity</td>
<td>Find statistics for similar entities (e.g., other schools, states, programs). Write down your best guess of how they should compare. How do the statistics actually compare? The statewide dropout rate for the prior year was 4.1%. The neighboring district reported 2.9%. Because your district is roughly between the two in demographics, you guess that your local rate should also be between theirs. 3.5% looks logical.</td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>4. Simple Math</td>
<td>Do some simple math with the statistic. Do the results make sense?</td>
<td>The technology report states that students average 2 hours a week on computers. You know the number of hours in a school day, the number of students, and the number of computers. Your simple calculations show that if every computer had a student on it every minute of the day, the average could only be 2 hours a week. Such efficient scheduling is impossible.</td>
</tr>
<tr>
<td>5. Calculate Counts</td>
<td>If the statistic is a percent, proportion, or ratio, calculate an actual count. Does this count make sense?</td>
<td>The report draft showed 12% of the students enrolled in AP English at the high school. That would be about 200 students. With only one AP English teacher, this doesn’t seem right.</td>
</tr>
<tr>
<td>6. Calculate Percents</td>
<td>If the statistic is a count, calculate a percent, proportion, or ratio. Does this calculation make sense?</td>
<td>The report showed 267 students eligible for a free lunch. That would be about 18% of the high school students. The high school must have at least 35% because it is one of your Title I schools.</td>
</tr>
<tr>
<td>7. Know the Source</td>
<td>Who is reporting the statistic? Are they the right person to do so? Are they the original source? Do you trust them?</td>
<td>The district’s music coordinator writes that 67% of college scholarship recipients were music students when in middle and high school. No source for the statistic is cited. You check and find that 67% of parents responding to a band booster survey said their child would receive some financial aid.</td>
</tr>
<tr>
<td>8. Independent Verification</td>
<td>Was the statistic independently verified?</td>
<td>The superintendent states that 82% of the district’s students passed the statewide math exam. The statistic is also reported by the state education agency and was calculated by the vendor for the assessment program.</td>
</tr>
<tr>
<td>Step</td>
<td>Description</td>
<td>Example</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>9. Graph Proportions</td>
<td>If there is a graph, are the scales and proportions appropriate?</td>
<td>A graph shows a dramatic increase in the number of students taking algebra. The y axis begins at zero and goes above the highest value shown. The ratio of the y to x axes is about 3 to 4. Everything appears to be done just like the text books suggest. So the impressive look of the graph is appropriate.</td>
</tr>
<tr>
<td>10. Details and Documentation</td>
<td>Are definitions, measures, limitations, samples, and other information provided for judging the validity of the statistic?</td>
<td>The evaluation that reported the algebra enrollments is accompanied by a technical report with the details.</td>
</tr>
<tr>
<td>11. Definitions and Periodicities</td>
<td>Do comparisons or changes reported use the same data points, definitions, periodicities, etc.?</td>
<td>Some problems are evident with the algebra enrollments. The current year is based upon beginning of the semester enrollment, but past years are counts of students earning credit. Past years include summer school, but the current year's summer is still in progress.</td>
</tr>
<tr>
<td>12. Stakes</td>
<td>What's at stake? How might the stakes have influenced the reporting of the statistic? How would competing perspectives have interpreted the statistic?</td>
<td>The high school is applying for a grant and must include achievement gains. The gains are impressive, but a change in school boundaries moved a large number of higher achieving students into the school last year. No adjustment for these students was made to verify that gains were made by the continuously enrolled students.</td>
</tr>
<tr>
<td>13. Gut Reaction</td>
<td>What's your gut reaction?</td>
<td>The district reports that dropouts have declined by 75% over the past five years. You haven’t noticed great changes, new programs, or any other intervention that could make such a huge difference. Reaction: You doubt this one.</td>
</tr>
</tbody>
</table>
The Data Quality Rating Scale

Use this to determine how good your data are.

Consumer Reports would want us to provide a rating system for data quality, so here’s one (Figure 2). Using the criteria of validity, accuracy, lateness, usefulness, and expense, an information source can be rated on this five-level scale. Try an area of data you are familiar with and apply the ratings. When I did this for the information systems I used to manage, the surprising winner was food service data. The loser? Discipline data. Make that undisciplined data.

Figure 2  Data Quality Rating Scale

<table>
<thead>
<tr>
<th>Information Source:</th>
<th>Source Type:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Validity</td>
<td>Accuracy</td>
</tr>
<tr>
<td>High Quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>There is a clear match between the data and the intended or primary use of the data. Appropriate comparisons can be made. Appropriate conclusions can be made.</td>
</tr>
<tr>
<td>Reduced Quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A relationship between the data and the intended or primary use of the data is assumed or is logical, but may not be well documented or proven.</td>
</tr>
<tr>
<td>Information Source:</td>
<td>Source Type:</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
</tr>
<tr>
<td></td>
<td><strong>Validity</strong></td>
</tr>
<tr>
<td>Low Quality</td>
<td>![Red Circle]</td>
</tr>
<tr>
<td>Poor Quality</td>
<td>![Red Circle]</td>
</tr>
<tr>
<td>Unknown Quality</td>
<td>![Question Mark]</td>
</tr>
</tbody>
</table>

Each of the rating components needs to be further detailed to ensure comparable ratings across raters. Accuracy is presented in Figure 3 as an example.
Figure 3  Accuracy Scale

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>The Data are Rated at the Level in Which ALL Conditions are Satisfied.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Quality</strong></td>
<td>81-85: A. Missing data are not well documented and impact use minimally. B. Data are certified by providers as accurate; problems are documented. C. Data standards and specifications are published and readily available to providers. 86-90: A. Missing data are well documented and impact use minimally. B. All data are certified by providers as accurate. C. Data standards and specifications are published and providers certify their compliance. 91-95: A. Missing data are well documented and do not impact use. B. All data have been verified as accurate by the collecting agency. C. Data standards and specifications are published and data are checked for compliance. 96-100: A. No data are missing. B. All data have been certified as accurate through audit or review. C. Data standards and specifications are published and data are in compliance.</td>
</tr>
<tr>
<td><strong>Reduced Quality</strong></td>
<td>61-65: A. Missing data limit use in at least one key area. B. Data problems are evident and limit use. C. Data standards and specifications are not relied upon. 66-70: A. Missing data limit use. B. Data problems are evident and may limit use. C. Data standards and specifications are not relied upon consistently. 71-75: A. Missing data are not documented and use is impacted. B. Data problems not documented and may limit use. C. Data standards and specifications do not provide adequate guidance to data providers. 76-80: A. Missing data are not well documented and use is impacted. B. Data problems are not fully documented and may limit use. C. Data standards and specifications are partially complete or in need of updating.</td>
</tr>
<tr>
<td>Low Quality</td>
<td>Data standards are weak or nonexistent. Poor controls are in place to ensure compliance.</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>41-45:</td>
<td>A. Most key data are missing. B. Data problems are pervasive and prevent use. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>46-50:</td>
<td>A. Substantial, key data are missing. B. Data problems are pervasive and prevent most use. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>51-55:</td>
<td>A. Missing data are prevalent enough to substantially limit use. B. Data problems are pervasive and substantially limit use. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>56-60:</td>
<td>A. Missing data are prevalent enough to require caution in use. B. Data problems are evident and substantially limit use. C. Data standards and specifications are not relied upon.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Poor Quality</th>
<th>Incorrect data, substantial missing data, or other problems are evident.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10:</td>
<td>A. Most data are missing. B. All data exhibit major problems. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>11-20:</td>
<td>A. Most data are missing. B. All data exhibit problems. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>21-30:</td>
<td>A. Most data are missing. B. Data problems are universal. C. Data standards and specifications are not available.</td>
</tr>
<tr>
<td>31-40:</td>
<td>A. Most data are missing. B. Data problems are substantial. C. Data standards and specifications are not available.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unknown Quality</th>
<th>Accuracy of the data is unknown or not documented.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Four Great Truths about Data Quality

Data quality is highest when…

1. The data providers know what’s expected.
2. The data providers use the data themselves for their own work.
3. Everyone, everywhere checks the data.
4. The data are available and used.

Part 1 of the Data Quality Series, The Data Quality Imperative, identified these four truths about data quality. They guided the design of the steps outlined in Attachment A: Data Quality, Best Practices for Local, State, and Federal Education Agencies.

Steps for Ensuring Data Quality

All the above is well and good—if not great in places. However, for those professionals on the line, designing and managing programs and information systems, there needs to be a users guide for data quality. There is.

Attachment A takes the principles and insights from this paper and translates them into the day-to-day activities that must be followed to achieve the highest level on the hierarchy.

A Final Note about Error

The hierarchy and the detailed steps do not deal completely with some of the nitty-gritty issues of data quality that are usually fretted over by information systems managers and data providers. Many of these fall into the general category of error. Error can be mistakes that result in bad data. Those have been addressed already. Error can also be measurement error (such as the standard error of measurement for an assessment) that keeps us from ever being 100% confident in our data.

Measurement errors are those imprecisions that result from our inability to be absolutely perfect in our measurements. One is the reliability of an instrument, test, or performance task (illustrated by a test-retest difference). Measurement errors can also be “intentional” as occurs when we round numbers or put values in ranges rather than use a more precise value. In research and evaluation situations, sampling error introduces its own limits on the reliability of the data. Measurement error should be recognized and acknowledged when data make their way to the reporting end of their life cycle.

Conclusion

Data quality is achievable if we establish the rules and follow them—all of us.

Attachment A follows.
**Data Quality**

Best Practice for Local, State, and Federal Education Agencies

Glynn D. Ligon, Ph.D. & Barbara S. Clements, Ph.D.

Data Quality:

Data quality is more than accuracy and reliability. High levels of data quality are achieved when information is valid for the use to which it is applied and when decision makers have confidence in the data and rely upon them.

Information Systems Architecture:

The foundation for data quality begins with a formal information systems architecture (ISA). The ISA is the metadata, hardware, software, and network standards, policies, governance, and requirements by which all information systems are built and managed. See the D3M Framework as described in Our Vision for D3M at http://www.espsg.com/resources.php.
### OVERVIEW

#### STEPS FOR ENSURING DATA QUALITY

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Are requirements known?</td>
<td>Validity</td>
<td>Accuracy</td>
<td>Timeliness</td>
<td>Editing</td>
<td>Reporting</td>
</tr>
<tr>
<td></td>
<td>Compare policy, regulation, rules, and procedures with the instructions given to data providers, collection forms, and code in software applications. M</td>
<td>Review design by peers, agencies, and staff. M</td>
<td>Provide training and certification for data providers. Train all new staff. M P</td>
<td>Use checklists and sign-offs for key steps. P</td>
<td>Run audit reports for review by experts with knowledge of reasonableness. M E</td>
<td>Fully disclose conditions affecting interpretation of the data. M P E</td>
</tr>
<tr>
<td></td>
<td>Include data providers and data processors in decisions to establish what is feasible. M</td>
<td>Precode all available data. Limit times data are entered. P</td>
<td>Precode all available data. Limit times data are entered. P</td>
<td>Run sample data and verify. P</td>
<td>Verify all calculations and conditional/business rules. M P E</td>
<td>Review data with providers and others with a stake in the results. M E</td>
</tr>
<tr>
<td></td>
<td>Follow an established change-management process. M P</td>
<td>Use most automated/validated level of data entry possible (e.g., selection from codes in an automated application vs. filling in fields). P</td>
<td>Provide documentation for data providers and data processors. M P</td>
<td>Ensure problems are identified, documented, corrected, and communicated back to the source of the problem or report. M P</td>
<td>Compare data to past runs, standards, or similar groups. M P E</td>
<td>Ensure analysis techniques meet the assumptions required for proper use. M E</td>
</tr>
<tr>
<td></td>
<td>Comply with professional standards for data collection, analysis, and reporting. M E</td>
<td>Use random checks during production. P</td>
<td>Provide immediate help for data providers. M P</td>
<td>Conduct on-site reviews during the process. M P</td>
<td>Check data exchanges, crosswalks, and translations for integrity. P</td>
<td>Present conclusions fairly within a context for interpretation. M E</td>
</tr>
<tr>
<td></td>
<td>Ensure people at all levels are knowledgeable, certified, trained, and competent for the tasks for which they are responsible. M E</td>
<td>Automate verification of entries at the earliest levels (e.g., upon keying Vs. from printed audit report). P</td>
<td>Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) M P</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Standards for Data Quality

**U.S. Department of Education**

**Persons Primarily Responsible for Data Quality During Each Step:**

- **M** = Manager of the program; designer of the collections; collector of the data; data steward
- **P** = Computer programmer; designer of the processing; processor of the data
- **E** = Evaluator; analyst; report writer

The provider of the data (e.g., school) is responsible for conscientiously following the prescribed process, reporting problems, and verifying the accuracy and completeness of all data submitted.
D3M Data Quality Process

1. Are requirements known?

   **Validity**
   - Compare policy, regulation, rules, and procedures with the instructions given to data providers, collection forms, and code in software applications. M
   - Include data providers and data processors in decisions to establish what is feasible. M
   - Follow an established change-management process. M P
   - Comply with professional standards for data collection, analysis, and reporting. M E
   - Ensure people at all levels are knowledgeable, certified, trained, competent, and energetic for the tasks for which they are responsible. M E

   **Timeliness**
   - Conduct stakeholder meetings.

Best Practice
DATA QUALITY STEP 1

- Build and maintain documentation for all programmatic and professional requirements.
- Conduct stakeholder meetings.
- Update files with copies of current/revised policies, regulations, rules, or procedures.
- Update references with current, new, or revised professional standards.
- Attend training or conference sponsored by controlling agency.
- Attend training or conference for professionals in this area.
- Conduct an annual personnel evaluation of all staff to ensure required competencies and performance are present. Follow-up with improvement plans and annual goals as appropriate.
- Create, review, or update job descriptions and competencies for each position related to this data area. Follow established procedures whenever a new employee is hired to ensure adequate qualifications.

- Communicate with programming staff to determine the extent and feasibility of all changes required.
- Communicate with data providers to determine the extent and feasibility of all changes required.
- Complete a “Program Change Request” for programmer to update code. Follow established change management process to implement changes.

Standards for Data Quality
U.S. Department of Education

Validity
- Attend training or conference sponsored by controlling agency.

Timeliness
- Attend training or conference for professionals in this area.

- Conduct an annual personnel evaluation of all staff to ensure required competencies and performance are present. Follow-up with improvement plans and annual goals as appropriate.
- Create, review, or update job descriptions and competencies for each position related to this data area. Follow established procedures whenever a new employee is hired to ensure adequate qualifications.
2. Is process well designed?

**Validity**
- Review design by peers, agencies, and staff. M

**Accuracy**
- Precode all available data. Limit times data are entered. P
- Use most automated/validated level of data entry possible (e.g., selection from codes in an automated application vs. filling in fields). P
- Use random checks during production. P
- Automate verification of entries at the earliest levels (e.g., upon keying Vs from printed audit report). P
- Run maintenance before all production. Verify off-hour maintenance and staff availability. P
- Ensure target dates are reasonable and clear. M

**Timeliness**
- Complete Step 1.
- Develop or review current process design aligned with current requirements from Step 1.
- Communicate with data providers to determine the extent and feasibility of all design changes required.
- Communicate with programming staff to determine the extent and feasibility of all design changes required.
- Review design with program management staff.
- Review design with officials in controlling agency.
- Review design with professionals in other districts.
- Incorporate validation processes into the design to ensure data integrity.
- Establish target dates for key actions and verify their reasonableness.
- Schedule maintenance of all hardware and checks of all other systems prior to key actions.
- Arrange for off-hour maintenance, on-call employees, and other back-up procedures during key activities.
- Arrange for a “hot back-up” site for all operations in the event of failure of primary systems.
- Access extant files to preprint all available data to eliminate entry errors by data providers.
- Automate data capture and incorporate edit checks and validations at the time of data entry.
- Schedule random checks during each phase of entry, processing, and production.
- Verify data at the earliest level; automate if possible.
- Integrate data entry/submission with transactions critical to the work of the data providers. Capture data directly from transactional systems. Maximize the re-use of data relied upon by the data providers for their own work. Create a dependence upon and ownership of the data by the data providers.
3. Is process well documented and communicated?

Provide training and certification for data providers. Train all new staff. MP

Provide documentation for data providers and data processors. MP

Provide immediate help for data providers. MP

Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

Accuracy

Timeliness

- Complete Step 1.
- Complete Step 2.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.

- Establish and follow proper district protocol for communicating with data providers to establish the authority and priority for providing data in this area.

- Establish and maintain a help system that quickly provides answers to data providers. Anticipate questions and information needs, and communicate with data providers. Include hardware, software, and network issues.

- Communicate with data providers whenever necessary to keep them informed and productive.

- Publish calendar of data activities.

- Provide training for data providers; document participation; measure skills and knowledge; issue certification of completion of training.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.

- Establish and follow proper district protocol for communicating with data providers to establish the authority and priority for providing data in this area.

- Establish and maintain a help system that quickly provides answers to data providers. Anticipate questions and information needs, and communicate with data providers. Include hardware, software, and network issues.

- Communicate with data providers whenever necessary to keep them informed and productive.

- Publish calendar of data activities.

- Provide training for data providers; document participation; measure skills and knowledge; issue certification of completion of training.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.

- Establish and follow proper district protocol for communicating with data providers to establish the authority and priority for providing data in this area.

- Establish and maintain a help system that quickly provides answers to data providers. Anticipate questions and information needs, and communicate with data providers. Include hardware, software, and network issues.

- Communicate with data providers whenever necessary to keep them informed and productive.

- Publish calendar of data activities.

- Provide training for data providers; document participation; measure skills and knowledge; issue certification of completion of training.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.

- Establish and follow proper district protocol for communicating with data providers to establish the authority and priority for providing data in this area.

- Establish and maintain a help system that quickly provides answers to data providers. Anticipate questions and information needs, and communicate with data providers. Include hardware, software, and network issues.

- Communicate with data providers whenever necessary to keep them informed and productive.

- Publish calendar of data activities.

- Provide training for data providers; document participation; measure skills and knowledge; issue certification of completion of training.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.

- Establish and follow proper district protocol for communicating with data providers to establish the authority and priority for providing data in this area.

- Establish and maintain a help system that quickly provides answers to data providers. Anticipate questions and information needs, and communicate with data providers. Include hardware, software, and network issues.

- Communicate with data providers whenever necessary to keep them informed and productive.

- Publish calendar of data activities.

- Provide training for data providers; document participation; measure skills and knowledge; issue certification of completion of training.

- Prepare, distribute, post, and maintain a comprehensive guide for data providers. Update guide as changes occur.

- Provide documentation for data providers and data processors. MP

- Provide immediate help for data providers. MP

- Ensure the physical and fiscal requirements are available (e.g., computer hardware, software, network, etc.) MP

- Follow the approved Information Systems Architecture to ensure compatibility with all standards, policies, and procedures.

- Document data standards in the Metadata Dictionary, including Level 1, 2, & 3 business rules. Document collections and repositories for authoritative data sources.

- Identify all data providers; maintain a record of experience and training.

- Identify new data providers as they are hired. Deliver training and certify their skills.

- Follow requirements for a data provider, e.g., education level, training, skills, experience, etc.

- Identify and keep current documentation of the hardware, software, network, and other resources required for each data provider.

- Inventory and follow-up on any deficiencies in the physical or fiscal requirements of data providers.

- Maintain a log of help requests, actions taken, and issues requiring follow-up for all data providers.
4. Is process well implemented?

- Use checklists and sign-offs for key steps. P
- Run sample data and verify. P
- Ensure problems are identified, documented, corrected, and communicated back to the source of the problem or report. M P
- Conduct on-site reviews during the process. M P

---

**Best Practice**

**DATA QUALITY STEP 4**

- Complete Step 1.
- Complete Step 2.
- Complete Step 3.

**Follow data-provider checklists for key actions and dates.**

**Follow data-processor checklists for key actions and dates.**

- Certify data completeness and accuracy by obtaining sign-offs if required by persons responsible before moving to the next step.
- Conduct on-site reviews to verify implementation.
- Run sample data through process to ensure everything is working according to specifications.
- Report, correct, and document discrepancies and problems.
- Communicate to data providers changes required for current process.
- Communicate changes required for Steps 1 - 4.

---

**Standards for Data Quality**

U.S. Department of Education

Send an “implementation report card” to data providers and their supervisors.
5. Are data verified and compared?

- **Editing**
  - Run audit reports for review by experts with knowledge of reasonableness. M E
  - Verify all calculations and conditional rules. M P E
  - Compare data to past runs, standards, or similar groups. M P E
  - Check data exchanges, crosswalks, and translations for integrity. P

- **Calculating**
  - Complete Step 1.
  - Complete Step 2.
  - Complete Step 3.
  - Complete Step 4.

- **Timeliness**
  - Build data submission file.

- **Document and log issues**
  - Review audit reports with experts with knowledge of reasonableness.
  - Review audit reports with experts with knowledge of requirements.

- **Verify data exchanges**
  - Crosswalks, and translations for integrity.

- **Verify all calculations**
  - Conditional rules, and Level 1, 2, & 3 business rules.

- **Certify completeness**
  - and accuracy and obtain sign-offs accepting the data.

- **Send a “data quality report card”**
  - to data providers and their supervisors.
6. Are data appropriately analyzed and reported?

- **Fully disclose conditions affecting interpretation of the data.** M E
- **Review data with providers and others with a stake in the results.** M E
- **Ensure analysis techniques meet the assumptions required for proper use.** M E
- **Present conclusions fairly within a context for interpretation.** M E
- **Publish technical reports or make available data files with detailed data for verification of analyses and statements.** M E
- **Protect the confidentiality rights of individuals (FERPA).** M E

---

**Best Practice**

**DATA QUALITY STEP 6**

- **Complete Step 1.**
- **Complete Step 2.**
- **Complete Step 3.**
- **Complete Step 4.**
- **Complete Step 5.**

- **STOP LOOK VERIFY**
- **Compile documentation for problems, special circumstances, and other factors impacting interpretation of the data.**
- **Create and publish reports with findings, conclusions, and disclosure of conditions impacting interpretations.**
- **Review final data and derived statistics with data providers, processors, program managers, knowledgeable peers, and others with a stake in the results.**
- **Verify that FERPA requirements are met in all reporting.**
- **Build final analysis file in compliance with specifications in the design.**
- **Conduct analyses in compliance with actual characteristics of and assumptions made about the nature of the data.**
- **Build data files or reports for public access to the level of data appropriate.**

- **STOP LOOK VERIFY**
- **Access log of problems and issues from Step 4.**
- **Access log of problems and issues from Step 5.**
- **Publish "Official Statistics" for use by all persons representing the organization and to form or compare to a baseline for describing trends.**
- **Publish guidelines for interpretation and use of the data, statistics, and reports.**
- **Manage access to all data, statistics, and reports through an education portal with directory services to determine authentication and authorization for all users.**

---

**Standards for Data Quality**

U.S. Department of Education
**About ESP Solutions Group**

ESP Solutions Group provides its clients with *Extraordinary Insight™* into PK-12 education data systems and psychometrics. Our team is comprised of industry experts who pioneered the concept of “data driven decision making” and now help optimize the management of our clients’ state and local education agencies.

ESP personnel have advised school districts, all 52 state education agencies, and the U.S. Department of Education on the practice of K-12 school data management. We are regarded as leading experts in understanding the data and technology implications of the *No Child Left Behind Act (NCLB), Education Data Exchange Network (EDEN), and the Schools Interoperability Framework (SIF).*

Dozens of education agencies have hired ESP to design and build their student record collection systems, federal reporting systems, student identifier systems, data dictionaries, evaluation/assessment programs, and data management/analysis systems.

To learn how ESP can give your agency *Extraordinary Insight* into your PK-12 education data, email info@espsg.com.

---

*This document is part of The Optimal Reference Guide Series, designed to help education data decision makers analyze, manage, and share data in the 21st Century.*