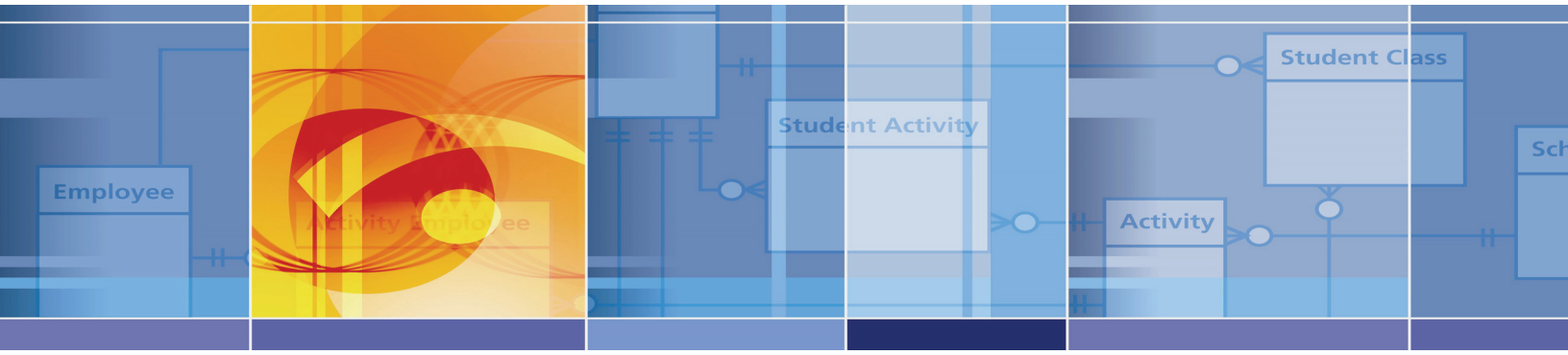


The Optimal Reference Book:

The Process for Ensuring Data Quality

Extraordinary insight™ into today's education information topics

ESP Solutions Group

Table of Contents

About ESP Solutions Group

ESP Solutions Group (www.espsolutionsgroup.com) is a PK-12 data consulting and technology firm specializing in education data systems and psychometrics. Our team is comprised of education experts who pioneered the concept of "data-driven decision making" (D3M) and now help optimize the management of our clients' state and local education agencies' information.

ESP is exclusively focused on PK-12 education. This is not a sideline business for our firm. We believe in what we do. We are former teachers, administrators, and district and state education agency personnel. ESP has a comprehensive view of the current state and future potential of the entire PK-12 data ecosystem. We understand how campus, district, state, and federal education technologies are related.

ESP personnel have advised local school districts, all 52 state education agencies and the extra-territorial jurisdictions, and the U.S. Department of Education on the practice of PK-12 school data management. We are nationally recognized as leading experts in understanding the data and technology implications of the No Child Left Behind Act (NCLB), Education Data Exchange Network (EDEN/ED*Facts)*, and Schools Interoperability Framework (SIF). We have also focused on the need for and requirements to implement PK-20 education data systems.

Since 1993, we have focused on delivering quality data into the hands of decision makers. We provide consulting services for information systems architecture planning and large-scale implementations. We also develop products and services for improved quality, collection, confidentiality, recovery, accessibility, and state and federal reporting. Our collective expertise is represented in our Optimal Reference Guides and Books. Recent timely topics such as growth models and action reports have joined our traditional papers on data warehouses and project management, and balance the thought-provoking "Data-Driven Decision Making 2016," "FERPA: Catch 1 through 22," and "Why Eva Baker Doesn't Seem to Understand Accountability." For our complete library of Optimal Reference Guides, Optimal Reference Books, and other education related resources, please visit www.espsolutionsgroup.com/resources.php.

Primarily, ESP brings the foremost education experts to the table for our clients. We have the knowledge, experience, and vision to insure our clients' long-term success. Our corporate headquarters are based in Austin, Texas. We have professionals located in seven other states around the nation both to represent various regions and to serve specific clients.

About the Author

Glynn D. Ligon, Ph.D., President and CEO

Dr. Ligon, the president and chief executive officer of ESP Solutions Group, is a nationally recognized expert on issues relating to student record collection and exchange, data quality, data reporting, and large-scale system design.

The National Center for Education Statistics, the U. S. Department of Education and over 25 state education agencies have consulted with Dr. Ligon on various areas of his expertise. He has a Ph.D. in Educational Psychology, Quantitative Methods from The University of Texas at Austin and is licensed to teach in the State of Texas.

Prior to starting ESP in 1993, Dr. Ligon directed the Austin (TX) Independent School District's information and technology organization. As the executive director of management information, he led the district's efforts in developing and reporting on district-wide program evaluations, many of which won national awards from the American Educational Research Association. Dr. Ligon was also a leader in the advent of SPEEDE/ExPRESS, the EDI standard for the exchange of electronic student transcripts.

From 1992 to 2000, he served as a member of the U.S. Department of Education's Planning and Evaluation Services Review Panel. Dr. Ligon's whitepapers; *A Technology Framework for NCLB Success* and *Steps for Ensuring Data Quality* are prominently featured within the U.S. Department of Education's 2005 National Education Technology Plan, meant to help motivate and incite technology-driven transformation in education.

At the beginning of his career, Dr. Ligon taught in predominantly Spanish-speaking schools near the Texas-Mexico border. He is an experienced evaluator of Title I, Migrant, compensatory education, and bilingual education programs.

Preface

One of the very best conference symposia I have seen (and over the years I have seen hundreds) was one on data quality at the 2009 NCES MIS Conference. The presentations were professional, as expected. The follow-up discussion was pointed and practical — and very insightful. This summary can't capture the banter and counterpoints to the audience's comments and questions. However, this paper can document the ideas contributed by the participants. Those ideas have significant value for all of us who work with education data and information systems designed to support data-driven decision making (D3M).

Jack Grayson, originator of the Malcolm Baldrige National Quality Award, has recently turned his attention to improving processes in education agencies. In 2002, I attended a six sigma briefing with Dr. Grayson, sat next to him all day, and discussed process improvement. Since that time, I have followed his quest to improve both processes and data quality for decision making in education. His organization, the American Productivity and Quality Center (APQC), has identified 20 "tools" for improvement. These are:

- After Action Review—originated and used by the U.S. Army to improve team performance by reflecting on action by comparing what is intended to what actually happened.
- Balanced Scorecard—a strategic management approach developed by Kaplan and Norton which measures processes, people, outcomes, and stakeholders.
- Malcolm Baldrige National Quality Award—continuous improvement criteria for leadership; strategic planning; student/stakeholder/market focus; measurement, analysis, and knowledge management; workforce focus; process management; organization performance results.
- Benchmarking—finding and adapting knowledge and best practices.
- Communities of Practice—people who share data, insight, experience, tips, information, knowledge, and advice about common interests or practices.
- Cycle Time—elapsed time of any process from beginning to end with the goal to reduce cycle time and increase data quality by reducing waste, reducing errors, and redesigning processes.
- Empowerment—involving employees.
- Evolving Technologies—typically refers to Web 2.0 technologies, e.g., blogs, P2P, RSS, wikis, social networking, etc.
- Hoshin Kanri Planning—direction setting management and daily management. Every employee understands long-range plans; all are working to a linked plan; process measures are continuously monitored.
- Implementation—knowing is not doing. Implementation is doing.
- Knowledge Management—a systematic process of identifying, capturing, organizing, sharing, transferring, adapting, and using data.
- Lean and Waste—a tool to remove anything that adds time or cost without adding value.
- Measures and Metrics—the yardsticks or standards by which to gauge performance; the numbers for gauging performance.
- Open Standards Benchmarking Collaborative (OSBC)—collects and analyzes benchmark process data in business, health care, and government.

- Education Process Management—now called North Star, APQC's education collaborative to improve processes.
- Quality Tools—process flow charting, benchmarking, six sigma, constraint analysis, affinity diagrams, balanced scorecards, etc.
- Six Sigma—systematic approach to reduce errors, decrease waste and costs, improve customer satisfactions, and increase revenues.
- Supply Chain—integrating function for linking major business functions and processes within and across entities into a cohesive and high performing model.
- Systems—processes have a defined purpose, goals, measures, and agenda.
- Transfer and Sharing—best practices must be shared across people and organizations.

These make a long list but they focus on a few simple concepts.

1. Processes are key.
2. Sharing what we know about processes is important.
3. Measuring processes is essential.

Let's compare what Dr. Grayson and his 20 tools or methodologies say to what our practitioners said in the symposium.

Our practitioners emphasized 1 (processes) and 2 (sharing); 3 (metrics) didn't get mentioned.

Our practitioners added one major emphasis of significance:

DATA PROVIDERS MUST KNOW WHAT'S EXPECTED OF THEM.

So, we all agree on the same core conclusion. To achieve data quality, we must manage the processes associated with our data very well.

This preface has two sections.

1. Practitioners' Insights
   Summary of comments from the presentations of the symposium members at the 2009 NCES MIS Conference in Seattle

2. The Steps for Ensuring Data Quality
   Process map detailing the steps for ensuring data quality throughout the year for an education agency

Section 1: Practitioners' Insights

The practitioners on the symposium's panel were:

- Sydney Fadaoff, Project Manager, Alaska Department of Education and Early Development
- Sonya Edwards, Education Administrator, Division of Data Management, California Department of Education
- Leigh Ann Grant-Engle, Data Manager, Missouri Department of Elementary and Secondary Education
- Mathew McIntyre, Database Administrator, Wyoming Department of Education
- Vince Meyer, Data Services Supervisor, Wyoming Department of Education

Their insights from the symposium were:

1. Automated collections with clear business rules, especially those converted or new systems that begin to gather record-level data rather than aggregate statistics, expose bad data rather than fix data problems.
2. The key conversion time to quality data is when the data providers become the data consumers. That's when they begin to rely upon the quality of their own data.
3. As data systems begin to be consolidated and the responsibility for the data falls to fewer individuals, the program staff have a tendency to relinquish responsibility to IT staff for data quality. Unfortunately, program staff are the people who in the past have been the ones who knew by sight if the data were right—if they made any sense.
4. The data governance culture must change in order to improve data quality.
5. Some data must be collected multiple times to ensure quality—to allow for verification. A single student identifier in a file cannot verify that this is the correct student record. Each file, or at least the source data that creates the file, must contain enough demographic or personal information for verification.
6. Silos can exist at the school, LEA, or SEA levels; and they must be managed at all levels. The SEA can eliminate all silos in its architecture, but the LEA may still maintain silos and redundancy; thus, all of the burden may not go away, even when the SEA has redesigned its systems.
7. Small schools and districts continue to be a challenge. The vendors do not support them as much. They are not as capable of meeting XML or other requirements. They cannot afford upgrades to applications. They may not even be motivated to be automated if automation is offered to them.
8. The more analytic power that people get, the more data they discover that they want. Each new report that someone sees creates the curiosity for another level of breakout or disaggregation.
9. When is a new requirement for data final? When should a school or district commit to programming for providing the data with the assurance that the

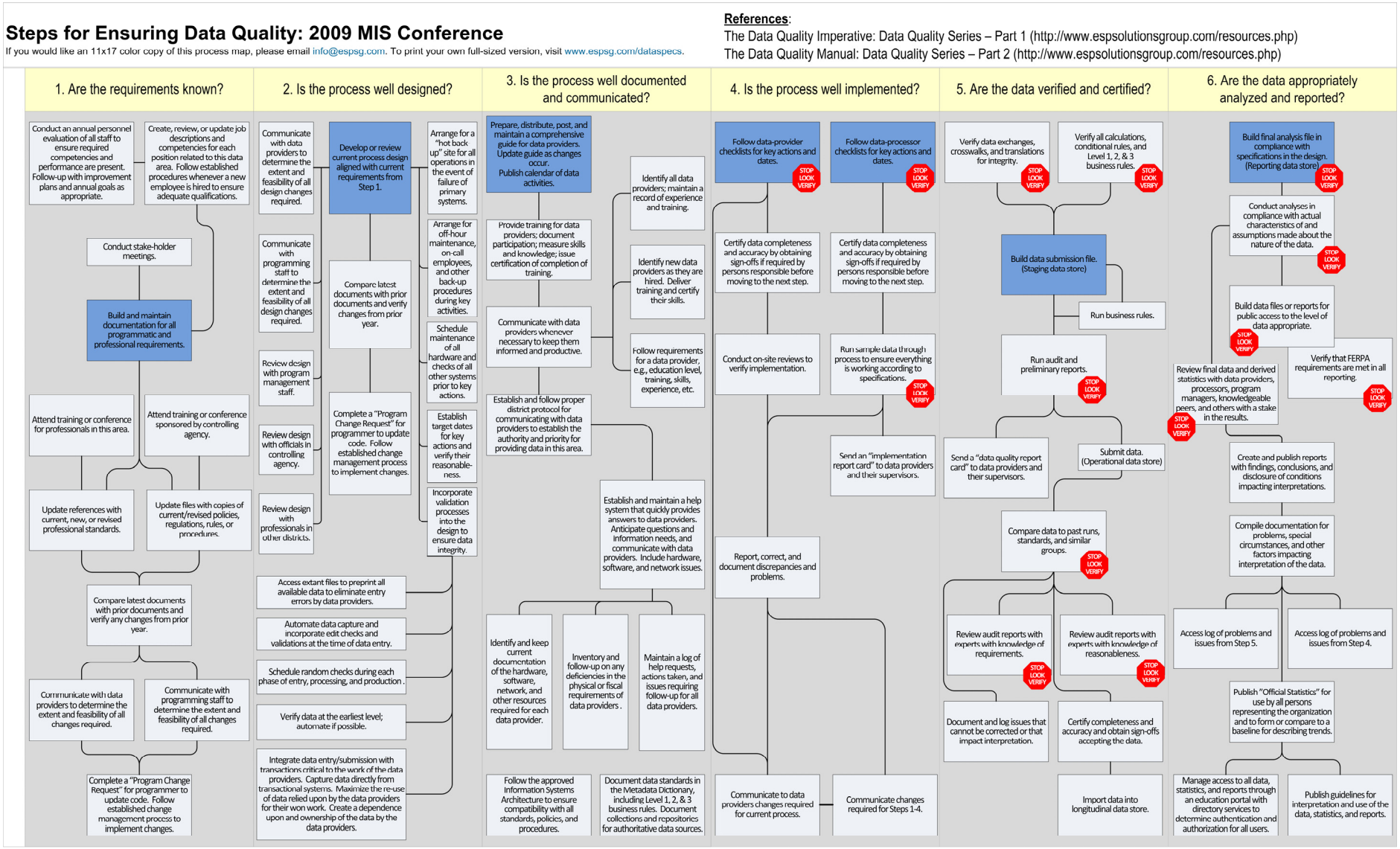
requirement will not change? Some SEAs have tried to give 6-12 months' lead time for changes to requirements; however, last-minute changes still occur. This remains a dilemma for districts and vendors.

10. Should schools and districts make changes to data online after files have been submitted or should changes be made in the source files and resubmitted? Everyone agrees that changes must be made to source files. However, minor changes that can be made to online files at the last minute are practical—as long as a process is in place to ensure they are also made to the source files.
11. How should vendors be engaged in data quality? This is a major, on-going issue.
12. LEA expertise may be more in the instructional IT arena rather than what is needed to support the SEA's processes. This is not surprising, but makes for a training and support challenge.
13. A comprehensive data access and management policy is needed to address not only FERPA and HIPPA, but also all the other issues related to how personally identifiable data are managed and used.

If you combine the hands-on knowledge from these practitioners with the best practices in our Data Quality Optimal Reference Book, you will be well on your way to better, more reliable data.

Section 2: Process Map – Steps for Assuring Data Quality

If you would like an 11x17 color copy of the process map below, please email info@espsg.com. To print your own full-sized version, visit www.espsg.com/dataspecs.



The Optimal Reference Guide:

The Data Quality Imperative, Data Quality Series – Part I

Foreword

By Barbara S. Clements, ESP Solutions Group

A few years ago, Glynn and I were asked to help create a Data Quality Manual for the Office of the Chief Information Officer of the U.S. Department of Education. Another contractor, called the Center for Data Quality (C4DQ), was also expected to help on the development of this manual. In fact, the C4DQ believed that this was their unique area of expertise and that they alone should develop the manual. When we pressed them for the areas they would cover in the manual, they described business rules and other data cleansing techniques. We convinced our client that this was only a small part of data quality and contributed much more detail for the manual. Much of that information is included in this series of papers on data quality.

ESP Insight

Our position has always been that data quality is dependent upon the initial entry of the data and the careful shepherding of the data throughout the system, including reporting.

Our position has always been that data quality is dependent upon the initial entry of the data and the careful shepherding of the data throughout the data system, including reporting. This paper makes the case for data quality as no one else has ever done – basically from soup to nuts. As I looked at the publications of various organizations and foundations that have focused on education data in recent years, I noticed that they were basically advocating for certain types of data that they think would be useful in the decision making process. Data quality is mentioned because of the perception that education data are "so bad." But these organizations can't wave a magic wand and get the data they desire, much less the quality of the data they desire. All components of the education system must be committed to providing the best quality data possible, and the participants in these components must be given the tools and training needed to improve the data. The second part of this series provides guidance on reducing typical errors in education data.

It is difficult for us to put all we've learned about data quality into a few documents. And we still continue to identify new areas of concern, error, and confusion. However, we are committed to sharing our experience and knowledge with educators and data managers. And hopefully we'll help educators make better decisions with better quality data.

Introduction

The secret to quality data is simple—get them right from the beginning.

Then keeping them right is a matter of effective data management processes.

What's the tell-tale sign that an agency does not have control over the quality of its data? If an education agency is spending time cleaning data, then the processes are wrong. Cleaning data is one of the least desirable tasks for an agency. The act of cleaning data means that someone earlier in the process did something wrong.

ESP Insight
The secret to data quality is simple—get them right from the beginning.

One of the biggest mistakes that an education agency makes when a new information system project is implemented is letting bad data get into the system. Bad data must be met at the door and turned away for the provider to correct. By the way, that door needs to be as early in the process as possible.

For this Optimal Reference Guide two-part series, best practices and principles for achieving data quality are detailed from start to finish. This is presented in these sections.

Part I: The Data Quality Imperative

1. Information Infrastructure – The components required to support data quality
2. Data Quality Boot Camp – Understanding the principles of data quality
3. Perspectives of Practitioners – How professionals who manage data view data quality
4. The Hierarchy of Data Quality – Getting to data-driven decision making

Part II: The Data Quality Manual (will be released in January, 2008)

5. Data Quality 101 – The "don'ts" to avoid messing up your data
6. Data Quality Checklist – Sensing the quality of data
7. The Data Quality Rating Scale – How good are your data?
8. Steps for Ensuring Data Quality – Start-to-finish steps to follow

ESP Insight
The tell-tale sign that an agency does not have control over data quality is the amount of time spent cleaning data.

These two papers present the clearest look into the dynamics of data quality yet developed by ESP's professionals. The reason—we've been helping education agencies improve their data quality since 1994. Before that, some of us were the ones sending in the data.

A simple test for the right attitude about data quality is how an agency reads these words.

CLEAN DATA

If you read clean as an action verb—that's trouble. If you read this as a statement of pride—there's hope.

The Imperative

Much talk buzzes around data quality. There's even a Campaign for Data Quality. This Optimal Reference Guide (ORG) describes quality data quality. That would be data quality of the highest order. That quintessential level of data quality is defined as:

ESP Insight
Data quality matters whether we are rating, funding, or selecting schools.

> **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 decision makers have confidence in the data and rely upon them.**

Samples of higher quality education data:

1. An official transcript certified by a high school registrar
2. A teacher's grade book for the end of a grading period—a week later
3. Teacher's certificate or license showing areas of certification or endorsement
4. Directory of sign-ons and passwords for a secure application
5. Payroll data—a month later
6. GIS file of addresses for enrolled student
7. Lunch eligibility status and meals served
8. Won-lost records of athletic teams in official events
9. School AYP status—after appeals

Samples of data that fall short of dependable quality:

1. Discipline data
2. Dollar amount of scholarships earned by graduating seniors
3. Student mobility rate
4. Student dropout rate
5. Instructional expenditures
6. Persistently dangerous schools
7. Hits on a school's website—what's a hit mean anyway?
8. Percent of high school graduates attending college—out of state
9. Number of ADHD students enrolled in kindergarten

When we rate schools and fund schools, data quality matters. When we describe schools out of curiosity (e.g., mobility rate, hits on a web site), data quality makes comparisons valid. When we select schools for our own kids to attend, softer data like anecdotes and opinions of trusted friends often trump the statistics—quality or otherwise. Whatever the purpose, we all want the best data possible.

Information Infrastructure – The components required to support data quality

Enhancing the Data Quality Campaign

I've spoken with the principals of the Data Quality Campaign recently about their work. DQC acknowledges that their name is somewhat of a misnomer. Their focus is on longitudinal data systems for state education agencies, but that doesn't inspire the buzz that data quality does. No problem—let's all take a wider perspective on data quality for awhile.

ESP Insight
Our focus on data quality goes far beyond longitudinal data systems. School funding and highly qualified teachers are two practical examples of the data quality imperative.

The 10 criteria DQC rated the states on are aging and limited. If anyone missed those ratings, DQC used Just for the Kids' (National Center for Educational Accountability) 10 components for a longitudinal data system to rate each state. Some states were miffed at their ratings. Ten criteria, self-reporting by the SEAs, public website reporting—SEAs have a point. From where did the criteria/components come? Are these really the components that impact data quality? The contents of the DQC components and fundamental concepts don't capture the sophistication of the solutions that an SEA must implement. Granted, they are all important requirements and should be in place, but they are not a representative subset of components upon which an SEA can evaluate its status.

DQC's focus on longitudinal data systems is too narrow. For example, data for school funding is another crucial data quality arena. Highly qualified teacher data? All states have those—to a degree.

Just for the Kids published virtually the same list of "9 components" in the 90's. Here's a new list to consider—the D3M Infrastructure Components. (The D3M Infrastructure was defined by the D3M Alliance—Microsoft, ESP Solutions Group, Edustructures, and Choice Solutions. (D3M = Data-Driven Decision Making). Attachment A is a representation of the high-level D3M Framework showing an SEA how the components work together.

The DQC list mixes subsystems with major systems and doesn't represent the relative significance of each either in terms of their importance and extensiveness. Their 10 official components are supplemented by their 7 fundamental concepts. So let's regroup everything into the D3M Infrastructure Components for comparison. See Table 1.

Table 1: D3M vs. DQC

D3M Infrastructure Component Description and Sample Contents		**DQC Component** (DQC's Numbering)	**DQC Fundamental Concept**	**Comments on DQC's Criteria for State Ratings**
Information Systems Architecture—the metadata, hardware, software, and network standards, policies, governance, and requirements by which all technology systems are built and managed	• Metadata Architecture, Standards, Business Rules • Systems Interoperability Standards • Policy, Governance, Stakeholder Guidance • Standards and Procedures that Guide All Other Components		• Data Architecture • Interoperability • Portability (Electronic Records Exchange)	Overall design and management not rated. Only 3 mechanical fundamental concepts—not even the most important ones. Don't forget policies, stakeholder buy-in, business rules, and many others.
Infrastructure—the physical hardware, software, network, and human resources required to support the technology systems	• Hardware • Systems Software • Application Software • Personal Productivity Software • LAN, WAN, and Internet Connectivity • Financial Underwriting • Human Resources • Business Continuation; and Disaster Prevention and Recovery			Infrastructure not rated. Must have the infrastructure perform perfectly to avoid introducing new data errors.

ESP Insight
The seven D3M infrastructure components capture the breadth of an agency's information system requirements.

D3M Infrastructure Component Description and Sample Contents		**DQC Component** (DQC's Numbering)	**DQC Fundamental Concept**	**Comments on DQC's Criteria for State Ratings**
Collections—the mechanisms for gathering data	• Data Specifications (Content), Business Rules, Periodicities • Data Quality, Certification of Submissions • Edit Reports and Help • Extensive Content about Students, Educators, Schools, Programs, Expenditures etc.	2. Student-Level Enrollment, Demographic, Program Information 4. Information on Untested Students and Reasons 6. Student Transcript Information 7. College Readiness Scores 8. Graduation and Dropout Data		DQC recommends some excellent contents, but again, far from all of the data required. Can't forget that data quality requires enforcing business rules upon entry, exchange, and reporting.
Data Stores—the centralized locations where data are located, managed, and accessed; includes a comprehensive data model	• Transactional • Staging • Operational • Longitudinal • Reporting • Back-Up	3. Match Test Records for Measurement of Academic Growth 10. State Audit System	Data Warehousing	Agreement—but much more content required, and the generic term "data warehousing" under-represents the sophistication required for managing an education agency's data.

D3M Infrastructure Component Description and Sample Contents		**DQC Component** (DQC's Numbering)	**DQC Fundamental Concept**	**Comments on DQC's Criteria for State Ratings**
Data-Driven Decision Support System—the way the data are provided to users for decision making, e.g., reports, queries, data files, etc.	• Action Reports, Standard Reports • Drill-Down Action Reports • *Ad Hoc* Query • Statistical Analysis (Research and Evaluation) • Profiles and Accountability Report Cards • Data Extracts			Interesting omission by DQC considering we have learned that data quality rises when the data are used and depended upon for decision making.
Portal—the system that authenticates and authorizes all users to provide appropriate access and security to all information	• Directory Services with Security for Authentication and Authorization • Individual Identifier and Locator System (Students, Teachers, All Others) • Collaboration Tools • Document and Resource Management • Notifications	1. Student Identifier 5. Teacher Identifier—to Match Students 9. PK-12 to Higher Education Records Matching	• Privacy Protection • Researcher Access	Portals are the modern way to manage all directory information and give users an interface that guides them through only the information they are authorized to see. Identifying all individuals allows for matching students, teachers, and interventions for analysis.
User Support—the system that trains, helps, and guides users to ensure efficient and proper use of the information	• Help and Support • Training and Professional Development • User Advisory Groups		Professional Development around Data Processes and Use	This has to be an official component, a reality, not just a concept.

An SEA doesn't simply have or not have these D3M components. The status of each is constantly changing—probably improving. For example, SEAs implementing new data warehouses already had some form of centralized data stores. All SEAs have some metadata standards, but everyone can improve—constantly.

What do the D3M components have to do with data quality? These components represent the entirety of an agency's information system. I guess that's the point—data quality is the result of everything within an information system working to perfection—not merely a set of simple rules to follow or data elements to collect. Or is it?

As detailed later on, there are 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.

ESP Insight
The "Four Great Truths" about data quality have held up after more than a decade of work with education agencies on quality issues.

How do these match with the D3M Infrastructure Components? See Table 2.

Table 2: The Truth about D3M Infrastructure Components

Truths about Data Quality	D3M Infrastructure Components						
	Information Systems Architecture	Infra-structure	Collections	Data Stores	Decision Support System	Portal	User Support
1. The data providers know what's expected.	**X**	X	X	X	X	X	**X**
2. The data providers use the data themselves for their own work.	X	X	X	X	**X**	X	**X**
3. Everyone, everywhere checks the data.	X	X	**X**	X	X	X	**X**
4. The data are available and used.	X	**X**	X	**X**	X	**X**	**X**

Read the chart as: Truth 1 is supported by all 7 components, but the Information Systems Architecture is the primary one.

Every component is supportive of all four great truths about data quality. The larger, green Xs indicate the primary contribution of each component. User Support is the common denominator for data quality across the board.

SEAs are making great progress improving data quality—but their challenge is immense. If every state had just the 10 DQC components in place today, data quality, timeliness of reporting, and appropriate use of longitudinal data might be good—but far short of great for all data. To make a real breakthrough moment happen, the SEAs need the 7 D3M Infrastructure Components functioning at a high level.

1. **Information Systems Architecture**—the metadata, hardware, software, and network standards, policies, governance, and requirements by which all technology systems are built and managed

2. **Infrastructure**—the physical hardware, software, network, and human resources required to support the technology systems

3. **Collections**—the mechanisms for gathering data

4. **Data Stores**—the centralized locations where data are located, managed, and accessed; includes a comprehensive data model

5. **Data-Driven Decision Support System**—the way the data are provided to users for decision making, e.g., reports, queries, data files, etc.

6. **Portal**—the system that authenticates and authorizes all users to provide appropriate access and security to all information

7. **User Support**—the system that trains, helps, and guides users to ensure efficient and proper use of the information

ESP Insight

Education agencies need all seven D3M infrastructure components to produce quality data from a comprehensive information system.

Data Quality Boot Camp – Understanding the principles of data quality

Ready to go through a boot camp for data quality? The basics of ensuring and maintaining quality data throughout an information system have been gleaned from our ESP experts and summarized below.

ESP Insight
The Data Quality Boot Camp is not merely platitudes. These are insights from across every state and a full range of information system architectures.

Data quality, the basics:

1. Get data right from the start.
2. Keep them right at every step.
3. Give people help to do this.

The next person in line can't fix the last person's errors as easily as that person can.

Poor data quality, the culprits:

1. Missing data
2. Incorrect data
3. Late data

Most vulnerable times for data:

1. Entry
2. Exchange

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.

Principles of Data Quality

Data quality abides by some well-tested principles. The fact that these are not widely known is a shame.

The *Expectation* Principle of Data Quality

- Data quality can only be achieved when the expectations are clear.
  Documentation of data definitions, codes, and business rules is essential. Metadata—be sure the data providers have been told.

The *Use* Principle of Data Quality

- Data quality matters when the data are used by the person collecting and reporting the data.
  The high school registrar is the law when it comes to official transcript data. The registrar must certify that the records are complete, accurate, and official, so nothing gets out without scrutiny.

The *Check Your Work* Principle of Data Quality

- Data quality requires all data handlers to check their own data.
  No one can spot errors and omissions in your data better than you. Don't pass along your errors and expect the next person to find and correct them.

The *Comparability* Principle of Data Quality

- Data quality matters when the data are compared.
  Is your school's attendance rate really lower than your rival's? Are you treating excused absences the same way?

ESP Insight
These Principles of Data Quality get translated into specific steps in Part II of this series on data quality.

The *Hierarchical Norm* Principle of Data Quality

- Each institution is a subject of a higher institution and an authority for a lower institution.
- Every data element an authority chooses to define must be defined the same by all lower institutions.

Institutional Hierarchy
US Department of Education
  State Education Agency
    Local Education Agency (District)
      School
        Employee

Notice that the individual tasked with providing the data is not an authority for the data.

The *Transformation* Principle of Data Quality

- A subject institution may define a data element differently from its authority only to the extent that the data element can be derived from or transformed into the precise definition of the higher authority.
  Keep more detail, use your own codes, but be sure you can transform it all to the required categories.

The *Transformation Burden* Principle of Data Quality

- Part A: The burden to transform is solely the burden of the subject institution.
- Part B: This burden compels the subject institution to comply with the standard of the authority.
  It's just easier to do it right the first time. Why have to transform your codes if you can use the standard ones from the beginning?

ESP Insight
Creativity, forgiveness, procrastination, and delegating upward are not principles in a quality data process.

The *Monkey on My Back* version of the Transformation Burden Principle of Data Quality

- Data Provider: I can get them to clean the data because they are the ones who want it anyway.
- Data Requestor: I'm the one who needs these data, so I have to clean them up if they won't.

    This is the root cause of so much pain. The requestor is the enabler. If rules are enforced from the beginning, data providers get the message that they can do it right now or do it again before the requestor will take it.

The *Invented Here* Principle of Data Quality

- Competes with the Transformation Burden Principle.
- As the local expert, I know how we should define our data.

    Not a team player, this know-it-all. The rules must be enforced even with the legendary staff members who have been around since the beginning of computer time.

The *Vendor Rules* Principle of Data Quality

- When we chose our vendor, we chose our data standards.

    No, no, no. Vendors want your business and your reference. Leverage that to get what you need.

The *Inertia* Principle of Data Quality

- If we change to use the authority's standard, we have to retrain everyone and reconfigure all our software.

    Yes, you do. Do it.

What does this mean for me?

- If you follow the authority's rules, burden is lower.
- If you change the rules, you have to re-work your data for reporting.

What does this mean for data quality?

- If people follow the rules, quality is higher.
- If people change the rules, quality is not achieved.

The unfortunate truth about reporting quality data:

- If you do something well the first time, people will not appreciate how difficult it is to do.

The redeeming factor:

- Getting data right from the start is difficult. However, providing clean, timely data is greatly appreciated by the collector.

Perspectives of Practitioners – How professionals who manage data view data quality

The following has been revised and enhanced since first being published as Data Quality: Earning the Confidence of Decision Makers, a paper presented at the annual meeting of the American Educational Research Association, April, 1996.

> **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.**

Professionals responsible for education data have long sought to provide timely and useful information to decision makers. Regardless of the evaluation model, research design, or statistical methodology employed, informing the decision making process with quality, reliable data is the basic goal. In the publications describing quality related to general information systems, the concept is narrowly interpreted to mean accurately and reliably processed data. This section ties together the foundations of data quality from the formal information systems literature with the practical aspects of data quality in the arena of public education decision making. A hierarchy of data quality has been developed to assist both the understanding of quality and the requirements for achieving quality. The hierarchy ranges from the availability of dysfunctional, bad data to the quality level of data-based decisions made with confidence.

ESP Insight
What are not universally sought are federally imposed standards for data and information systems.

Background

Data quality is essential to successful research, evaluation, and statistical efforts in public schools. As statewide accountability systems that rely upon large data bases grow, concern follows about the data quality within those emerging state-level databases. As states and the Federal government move toward establishing data stores to make information available electronically to everyone, questions are raised about the quality of the data collected and stored.

What are not universally sought are federally imposed standards for data and information systems. There is broad support for voluntary standards which states and local school districts can adopt. What is needed first is a way to know when quality data are available and when caution should be exercised. All this must be accomplished within the context of the ever-changing world of information technology. Several of our papers provide this context—past, present, and future.

- ***New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education***
  Published by the National Center for Education Statistics, U.S. Department of Education, 1996
- ***How Education Data Fared in the Last Decade***
  ESP Optimal Reference Guide, August 2006
- ***Data Driven Decision Making 2016***
  ESP Optimal Reference Guide, November 2006

Decision makers at all levels are relying upon data to inform, justify, and defend their positions on important issues. What are the key criteria on which to determine data quality? Is there a logical sequence to the processes for ensuring quality in information systems?

ESP Insight
The concern for data quality is a sign of maturity in the field, an increasing sophistication by the audiences who use education data.

The concern for data quality is somewhat different than the slowly emerging interest in education data that has grown for decades. The concern for data quality is a sign of maturity in the field, an increasing sophistication by the audiences who use education data. In other words, first we asked "Are our students learning?" Then we had to ask "What are the education indicators that we should be monitoring?" Finally, we are asking "Now that we have some indicators, do we trust them?" Several of our papers provide context around education indicators.

- ***What Dow-Jones Can Teach Us: Standardized Education Statistics and Indicators***
  G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1993
- ***A Dow Jones Index for Educators***
  G. Ligon, The School Administrator, December, 1993
- ***Actions Speak Louder than Data***
  ESP Optimal Reference Guide, March 2007
- ***From Information to Insight—the point of indicators***
  ESP Optimal Reference Guide, April 2007

An easy point in time to mark is the release of the "Nation at Risk" report. Much reform in education followed, including expansion of accountability systems within states. The search heated up for the true, reliable indicators of quality in education. Another major event was the passage of the 1988 Hawkins Stafford Education Amendments that called for improving the quality of the nation's education data. From that legislation, the National Forum for Education Statistics was begun, and from that group has followed a continuing focus on data quality issues. The Forum, sponsored by the National Center for Education Statistics, which is part of the Institute for Education Sciences, is made up of state education agency representatives and local education agency staff.

Then in 2001, everything was ratcheted up several notches with the passage of the No Child Left Behind Act. SEAs suddenly began taking the data challenges presented by accountability mandates very seriously.

There are multiple perspectives, each with its own reality of data quality. These are:

- Decision Makers (parents, teachers, counselors, principals, school board members, legislators, governors)
- Program Managers (directors, supervisors)
- General Audiences (news media, taxpayers, businesses)
- Data Collectors and Providers (clerks, teachers, counselors, program managers)
- Analysts (evaluators, researchers)

Individuals may occupy more than one of these groups simultaneously.

At the risk of over simplifying, the primary perspective of each group may be described as:

Decision Makers:
"Do I have confidence in the data and trust in the person providing them?"

Program Managers:
"Do the data fairly represent what we have accomplished?"

General Audiences:
"Did I learn something that appears to be true and useful, or at least interesting?"

Data Collectors and Providers:
"Did the data get collected and reported completely and in a timely manner?"

Evaluators, Researchers, Analysts:
"Are the data adequate to support the analyses, results, and interpretations from them?"

ESP Insight
In the end, the audiences (e.g., program managers, decision makers, and general audiences) give the ultimate judgment of quality when they use, ignore, or disregard the data.

The burden for data quality traditionally falls to the data collectors and providers. Who else would be in a better position to monitor and judge data quality? However, in the end, the audiences (e.g., program managers, decision makers, and general audiences) give the ultimate judgment of quality when they use, ignore, or disregard the data. Our conclusion? *The highest level of data quality is 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.*

The Pursuit of a Definition of Data Quality

Years ago, Robert Friedman, formerly the director of the Florida Information Resource Network (FIRN), Arkansas's statewide network, and the California Student Information System (CSIS), called me and asked for references related to data quality. The issue had arisen as the new statewide education information system for Arkansas was being developed. There were few references available, none satisfactory. I began documenting anecdotes, experiences, and insights provided by individuals within the education research, evaluation, and information systems areas to search for "truths." Three years after Friedman's inquiry, I responded with the following insights.

Several ideas were consistently referenced by individuals concerned with data quality.

1. Accuracy
Technical staff mention reliability and accuracy. This is consistent with the published literature in the information systems area. Accuracy, accuracy, accuracy—defined as do exactly what we are told, over and over. Not all information specialists limit themselves to the mechanical aspects of accuracy; however, because they may not be content or process specialists in the areas they serve, their focus is rightfully on delivering exactly what was requested. After all, that is what the computer does for them.

Quality data in, quality data out.

2. Validity
However, programmatic staff point out that data must be consistent with the construct being described (i.e., validity). If their program is aimed at delivering counseling support, then a more direct measure of outcomes than an achievement assessment is desired.

Valid data are quality data.

ESP Insight
A key element frequently cited as basic for achieving quality is the reliance upon and use of the data by the persons responsible for collecting and reporting them. This may be the most important truism in this paper.

3. Investment
A key element frequently cited as basic for achieving quality is the reliance upon and use of the data by the persons responsible for collecting and reporting them. School clerks who never receive feedback or see reports using the discipline data they enter into a computer screen have little investment in the data. School clerks who enter purchasing information into an automated system that tracks accounts and balances have a double investment. They save time when the numbers add up, and they receive praise or complaints if they do not. Whoever is responsible for collecting, entering, or reporting data needs to have a natural accountability relationship with those data. The data providers should experience the consequences of the quality of the data they report.

This may be the most important truism in this paper:

The user of data is the best recorder of data.

4. Certification
Typically, organizations have a set of "official" statistics that are used, regardless of their quality, for determining decisions such as funds allocation or tracking changes over time. These official statistics are needed to provide some base for planning, and the decision makers are challenged to guess how close they are.

Organizations should certify a set of official statistics.

5. Publication
Public reporting or widespread review is a common action cited in the evolution of an information system toward quality.

In every state that has instituted a statewide accountability system, there are stories of the poor quality of the data in the first year. Depending upon the complexity of the system and the sanctions imposed, (either money or reputation) subsequent improvements in data quality were seen.

The most practical and easily achieved action for impacting data quality is:

Publish the data.

6. Trust
Decision makers refer to the trust and confidence they must have in both the data and the individuals providing the data.

Trust is a crucial component of the working relationship between decision makers and staff within an organization. That trust must be present for data to be convincing. Consultants are used at times to provide that trust and confidence. Decision makers often do not have the time nor the expertise to analyze data. They rely upon someone else's recommendation. Data should be presented by an individual in whom the decision makers have confidence and trust.

ESP Insight
Trust must be present for data to be convincing.

Trust the messenger.

These six statements faithfully summarize the insights of professionals who have struggled with data quality within their information systems. They address processes that contribute toward achieving data quality—the dynamics influencing quality within an information system. They do not yet clearly indicate how successful the organization has been in achieving quality. To make that connection, the following hierarchy was developed.

A Hierarchy of Data Quality – Getting to Data-Driven Decision Making

This hierarchy of data quality was designed in the 90's to describe how quality develops and can be achieved. This section details the components and levels within this hierarchy. This schema is to be regarded as fluid within an organization. Some areas of information, such as student demographics, may be more advanced than others, such as performance assessments. Some performance assessments may be more advanced than others.

ESP Insight
The highest level of quality is achieved when data-based decisions are made with confidence.

The highest level of quality is achieved when data-based decisions are made with confidence. Therefore, several components of quality must be present, i.e., available data, decisions based upon those data, and confidence by the decision maker. Ultimately, quality data serve their intended purpose when the decision maker has the trust to use them with confidence. The traditional virtues of quality (e.g., reliability and validity) form the basis for that trust, but do not ensure it. Accuracy is the traditional characteristic defined within formal information systems architecture. Accuracy begs the question of whether or not the data are worthy of use.

From the observations of organizational quests for quality information systems, the concept of official data has been described. Data are official if they are designated as the data to be used for official purposes, e.g., reporting or calculation of formulas such as for funding schools and programs. At the earliest stages of information systems, the characteristic of being available is the only claim to quality that some data have. The level at the base of the hierarchy is characterized by no data being available.

Attachment B summarizes and represents the hierarchy in visual form.

Examples are provided below to illustrate each level. As you will notice, most of these are from the 80's and 90's when I was managing information systems in a local school district. I was more comfortable using these examples from my own work than more recent ones from our ESP client engagements.

Bad Data

-1.1 Invalid

Bad data can be worse than no data at all. At least with no data, decision makers rely upon other insights or opinions they trust. With bad data, decision makers can be misled. Bad data can be right or wrong, so the actual impact on a decision's outcome may not always be negative. Bad data can result from someone's not understanding why two numbers should not be compared or from errors and inconsistencies throughout the reporting process. The definition of bad data is that they are either:

- Poorly standardized in their definition or collection to the extent that they should be considered unusable, or
- inaccurate, incorrect, unreliable.

An example of bad data occurred when a local high school failed to note that the achievement test booklets being used were in two forms. The instructions were to ensure that each student received the same form of the exam for each subtest. However, the booklets were randomly distributed each day of the testing, resulting in a mixture of subtest scores that were either accurate (if the student took the form indicated on the answer document) or chance level (if the form and answer document codes were mismatched). This high school was impacted at the time by cross-town bussing that created a very diverse student population of high and low achievers. From our previous analyses, we also knew that an individual student's scores across subtests could validly range plus or minus 45 percentile points. Simple solutions to interpreting the results were not available. (***Empty Bubbles: What Test Form Did They Take?*** D. Doss and G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1985.)

ESP Insight
Decision makers who want to use data or want to support a decision they need to make are vulnerable to grasping for any and all available data—without full knowledge of their quality.

Carolyn Folke, Information Systems Director for the Wisconsin Department of Education, contributed the notion that the hierarchy needed to reflect the negative influence of bad data. In her experience, decision makers who want to use data or want to support a decision they need to make are vulnerable to grasping for any and all available data—without full knowledge of their quality. The message here is look into data quality rather than assume that any available data are better than none.

None

0.0 Unavailable

Before "A Nation at Risk," before automated scheduling and grade reporting systems, and before the availability of high-speed computers, often there were no data at all related to a decision. So, this is really the starting point for the hierarchy.

When a local school district began reporting failure rates for secondary students under the Texas No Pass/No Play Law, one school board member asked for the same data for elementary students. The board member was surprised to hear that, because elementary grade reporting was not automated, there were no data available. (After a long and painful process to collect elementary grade data, the board member was not pleased to learn that very few elementary students ever receive a failing grade and that fewer fail in the lower achieving schools than fail in the higher achieving schools.) (***No Pass - No Play: Impact on Failures, Dropouts, and Course Enrollments,*** G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1988.)

When no data are available, the options are typically obvious—collect some or go ahead and make a decision based upon opinion or previous experience.

However, there is another option used by agencies involved in very large-scale data collections. The Bureau of the Census and the National Center for Education Statistics both employ decision rules to impute data in the absence of reported numbers. Missing cells in tables can be filled with imputed numbers using trends, averages, or more sophisticated prediction analyses. Decision makers may perform their own informal imputations in the absence of data.

Available

1.1 Inconsistent Forms of Measurement

Poor data come from inconsistencies in the ways in which outcomes or processes are measured. These inconsistencies arise from use of nonparallel forms, lack of standardized procedures, or basic differences in definitions. The result is data that are not comparable.

ESP Insight
Equating each state's performance levels with those of NAEP is a popular method for judging the difficulty of assessments across states.

In 1991, we studied student mobility and discovered that not only did districts across the nation define mobility differently, but they also calculated their rates using different formulas. From 93 responses to our survey, we documented their rates and formulas, and then applied them to the student demographics of Austin. Austin's "mobility" rate ranged from 8% to 45%, our "turbulence" rate ranged from 10% to 117%, and our "stability" rate ranged from 64% to 85%. The nation was not ready to begin comparing published mobility rates across school districts. (***Student Mobility Rates: A Moving Target,*** G. Ligon and V. Paredes, Presented at the American Educational Research Association Annual Meeting, 1992.)

A future example of this level of data quality may come from changes in the legislation specifying the nature of evaluation for Title I Programs. For years, every program reported achievement gains in normal curve equivalent units. Current legislation requires each state to establish an accountability measure and reporting system. Equating each state's performance levels with those of NAEP is a popular method for judging the difficulty of assessments across states.

Full time equivalents and head counts, duplicated and unduplicated counts, average daily attendance and average daily membership are all examples of how state accountability systems must align the way schools maintain their records. Who is not familiar with the "problem" of whether to count parents in a PTA meeting as one attendee each or as two if they have two students in the school?

1.2 Data Collected by Some at Some Times

Incomplete data are difficult to interpret.

In 1994, the *Austin American Statesman* published an article about the use of medications for ADD/ADHD students in the public schools. The headline and point of the story was that usage was much lower than had been previously reported. The person quoted was not a school district employee and the nature of some of the statistics caused further curiosity. So, I called the reporter, who said he had not talked to the District's Health Supervisor and that the facts came from a graduate student's paper. Checking with the Health Supervisor showed that only about half the schools had participated in the survey, some of those with the highest levels of use did not participate, the reporter used the entire District's membership as the denominator, and the actual usage rate was

probably at least twice what had been reported. The reporter's response: "I just reported what she told me."

1.3 Data Combined, Aggregated, Analyzed, Summarized

The highest level of "available data" is achieved when data are summarized in some fashion that creates interesting and useful information. At this point in the hierarchy, the data begin to take on a usefulness that can contribute to a cycle of improved quality. At this point, audiences are able to start the process of asking follow-up questions. The quality of the data becomes an issue when someone begins to use summary statistics.

ESP Insight
Data become information when decision-makers can understand them.

One of the most dramatic responses to data I recall was when we first calculated and released the numbers and percentages of overage students, those whose age was at least one year over that of their classmates. Schools have always had students' ages in the records. Reality was that no one knew that by the time students reached grade 5 in Austin, one out of three was overage. In at least one elementary school over 60% of the fifth graders were old enough to be in middle school. (The number of elementary retention's began to fall until the rate in the 90's was about one fifth of the rate in the 80's.) (***Do We Fail Those We Fail?***, N. Schuyler and G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1984; ***Promotion or Retention,*** Southwest Educational Research Association Monograph, G. Ligon, Editor, 1991.)

When relatively unreliable data are combined, aggregated, analyzed, and summarized, a major transformation can begin. Decision makers can now apply common sense to the information. Data providers now can see consequences from the data they report. This is an important threshold for data quality. In countless conversations with information systems managers and public school evaluators, a consistent theme is that when people start to see their data reported in public and made available for decision making, they begin to focus energies on what those data mean for them and their school/program.

Texas schools began reporting financial data through PEIMS (Public Education Information Management System) in the 1980's. The first data submissions were published as tables, and for the first time it was simple to compare expenditures in specific areas across schools and districts. Immediately, a multi-year process began to bring districts more in line with the State's accounting standards and to ensure better consistency in the matching of expenditures to those categories. When districts reported no expenditures in some required categories and others reported unrealistically high amounts, the lack of data quality was evident. The persistent lack of consistency across districts prompted the Texas Legislature in 2006 to fund a new study and development of a more standardized financial reporting process.

DATA BECOME INFORMATION. *Around this point in the hierarchy, data become information. The individual data elements are inherently less useful to decision makers than are aggregated and summarized statistics. From this point on in the hierarchy, basic data elements are joined by calculated elements that function as indicators of performance.*

Official

2.1 Periodicity Established for Collection and Reporting

Periodicity is the regularly occurring interval for the collection and reporting of data. An established periodicity is essential for longitudinal comparisons. For valid comparisons across schools, districts, and states, the same period of time must be represented in everyone's data.

ESP Insight

Periodicity—an agency must manage the periodicity of its data to understand what is available when.

The National Center for Education Statistics (NCES) has established an annual periodicity set around October 1 as the official date for states to report their student membership. Reality is that each state has its own funding formulas and laws that determine exactly when membership is counted, and most do not conduct another count around October 1 for Federal reporting.

I was called on the carpet by my local superintendent once because a school board member had used different dropout rates than he was using in speeches during a bond election. He explained very directly that "Every organization has a periodicity for their official statistics." That of course is how they avoid simultaneous speeches using different statistics. After working hard with the staff to publish a calendar of our official statistics, I discovered that very few districts at the time had such a schedule. (***Periodicity of Collecting and Reporting AISD's Official Statistics***, G. Ligon et al., Austin ISD Publication Number 92.M02, November, 1992.)

2.2 Official Designation of Data for Decision Making

Finally, official statistics make their way into the hierarchy. The key here is that "official" does not necessarily guarantee quality. Official means that everyone agrees that these are the statistics that they will use. This is a key milestone, because this designation contributes to the priority and attention devoted to these official statistics. This in turn can contribute to on-going or future quality.

Sometimes politimetrics turn out to be better than legacy statistical processes. Every year, our Management Information Department's Office of Student Records issued its student enrollment projection. The preliminary projection was ready in January for review and a final projection for budgeting was ready by March. Here is another example of how the presence of a bond election can influence the behavior of superintendents and school board members. The superintendent gave a speech to the Chamber of Commerce using the preliminary projection. Then our office sent him the final projection. He was not happy with the increase of about 500 in the projection. He believed that created a credibility gap between the figures used in campaigning for the bonds and the budgeting process. So, the preliminary projection, for the first time in history, became the final, "official" projection. The bonds passed, the next year's enrollment was only a few students off of the "official" projection, the School Board was impressed with the accuracy of the projection, and Austin began a series of four years when all the projection formulas were useless during the oil and real estate bust of the late 80's. The next time the "official" projection was close was when a member of the school board insisted that the

district cut 600 students from its projection in order to avoid having to budget resources to serve them.

THE RIGHT DATA MUST BE USED. *At this point, the qualities of accuracy and reliability are required. Moreover, the best data are not quality data if they are not the right data for the job.*

2.3 Accuracy Required for Use in Decision Making

With the official designation of statistics, either by default or intent, their use increases. Now the feedback loop takes over to motivate increased accuracy. The decision makers and the persons held accountable for the numbers now require that the data be accurate.

ESP Insight
The analyses must be appropriate to the question being addressed. The handy process available with an analysis tool may not meet the required assumptions for your data.

When we began publishing six-week dropout statistics for our secondary schools, the principals started to pay attention to the numbers. They had requested such frequent status reports so the end-of-the-year numbers would not be a surprise, and so they could react if necessary before the school year was too far along. Quickly, they requested to know the names of the students that we were counting as dropouts, so verification that they had actually dropped out could be made. Having frequent reports tied directly to individual student names improved the quality of the dropout data across the schools.

THE RIGHT ANALYSES MUST BE RUN. *The quality of data is high at this point, and the decision maker is relying upon analyses conducted using those data. The analyses must be appropriate to the question being addressed.*

*A caution to data providers and audiences: There are times when data quality is questioned, but the confusing nature of the data comes from explainable anomalies rather than errors. We should not be too quick to assume errors when strange results arise. For example, a district's overall average test score can decline even when all subgroup averages rise; students can make real gains on performance measures while falling farther behind grade level; schools can fail to gain on a state's assessment, but be improving. (**Anomalies in Achievement Test Scores: What Goes Up Also Goes Down,*** G. Ligon, Presented at the American Educational Research Association Annual Meeting, 1987.)

Valid

3.1 Accurate Data Consistent with Definitions

Trained researchers are taught early to define operationally all terms as a control in any experiment. Every organization should establish a standard data dictionary for all of its data files. The data dictionary provides a definition, formulas for calculations, code sets, field characteristics, the periodicity for collection and reporting, and other important descriptions. Using a common data dictionary provides the organization the benefits of efficiency by avoiding redundancy in the collection of data elements. Another important benefit is the ability to share data across departmental data files. (***Periodicity™ User Guide***, Evaluation Software Publishing, Austin, Texas, 1996.)

The classic example of careless attention to definitions and formulas is *Parade Magazine*'s proclamation that an Orangeburg, South Carolina, high school

reduced its dropout rate from 40% to less than 2% annually. Those of us who had been evaluating dropout-prevention programs and calculating dropout rates for a number of years became very suspicious. When newspapers around the nation printed the story that the dropout rate in West Virginia fell 30% in one year after the passage of a law denying driver's licenses to dropouts, we were again skeptical. Both these claims had a basis in real numbers, but each is an example of bad data.

ESP Insight
Authentic assessments have failed the bias test and remain useful for formative but not accountability processes.

The *Parade Magazine* reporter compared a four-year, longitudinal rate to a single-year rate for the Orangeburg high school. The newspaper reporter compared West Virginia's preliminary dropout count to the previous year's final dropout count. (The West Virginia state education agency later reported a change from 17.4% to about 16%.) (***Making Dropout Rates Comparable: An Analysis of Definitions and Formulas,*** G. Ligon, D. Wilkinson, and B. Stewart, Presented at The American Educational Research Association Annual Meeting, 1990.)

3.2 Reliable Data Independent of the Collector

Reliability is achieved if the data would be the same regardless of who collected them.

What better example is available than the bias in teacher evaluations? When Texas implemented a career ladder for teachers, we had to certify those eligible based upon their annual evaluations. The school board determined that they were going to spend only the money provided by the State for career ladder bonuses, so that set the maximum number of teachers who could be placed on the career ladder. Our task was to rank all the eligible teachers and select the "best." Knowing there was likely to be rater bias, we calculated a Z score for each teacher based upon all the ratings given by each evaluator. Then the Z scores were ranked across the entire district. The adjustments based upon rater bias were so large, that near perfect ratings given by a very easy evaluator could be ranked below much lower ratings given by a very tough evaluator. The control was that the teachers' rankings within each rater's group were the same.

Everything was fine until a school board member got a call from his child's teacher. She was her school's teacher-of-the-year candidate but was ranked by her principal in the bottom half of her school, and thus left off the career ladder. The end of the story is that the school board approved enough additional local money to fund career ladder status for every teacher who met the minimum state requirements, and we were scorned for ever having thought we could or should adjust for the bias in the ratings. (***Adjusting for Rater Bias in Teacher Evaluations: Political and Technical Realities,*** G. Ligon and J. Ellis, Presented at the American Educational Research Association Annual Meeting, 1986.)

3.3 Valid Data Consistent with the Construct Being Measured

The test of validity is often whether a reasonable person accountable for an outcome agrees that the data being collected represent a true measure of that outcome. Validity is the word for which every trained researcher looks. Validity assumes both accuracy and reliability. Critically, valid data are consistent with the construct being described. Another perspective on this is that valid data are those that are actually related to the decision being made.

> The local school board in discussing secondary class sizes looked at the ratio of students to teachers in grades 7 through 12 and concluded that they were fairly even. Later they remembered that junior high teachers had been given a second planning period during the day, so their actual class sizes were much higher. Then they moved on to focus on the large discrepancies between class sizes within subject areas to discover that basic required English and mathematics classes can be efficiently scheduled and are large compared to electives and higher level courses. In the end, the school board members became more understanding of which data are valid for use dependent upon the questions they are asking.

ESP Insight
Quality data must be interpretable beyond the local context.

Quality

4.1 Comparable Data: Interpretable Beyond the Local Context

Quality is defined here beyond the psychometric and statistical concepts of reliability and validity. Quality is defined by use. Quality data are those that function to inform decision making. For this function, the first criterion is:

> Quality data must be interpretable beyond the local context. There must be a broad base of comparable data that can be used to judge the relative status of local data. We can recognize that there are some decisions that do not necessitate comparisons, but in most instances a larger context is helpful. Each time I read this criterion, I rethink it. However, it is still in the hierarchy because decisions made within the broadest context are the best informed decisions. Knowing what others are doing, how other districts are performing does not have to determine our decisions, but such knowledge ensures that we are aware of other options and other experiences.

ESP Insight
Data-based decisions must be made with confidence, at least confidence in the data.

> Most states and districts have struggled with defining and reporting their dropout rates. Despite the lofty goal often embraced of having 100% of our students graduate, there is still the need for comparison data to help interpret current levels of attrition. When we compared Austin's dropout rate to published rates across the nation, we found that the various formulas used by others produced a range of rates for Austin from 11% to 32%. Our best comparisons were across time, within Austin, where we had control over the process used to calculate comparable rates. (***Making Dropout Rates Comparable: An Analysis of Definitions and Formulas,*** G. Ligon, D. Wilkinson, and B. Stewart, Presented at The American Educational Research Association Annual Meeting, 1990.)

4.2 Data-Based Decisions Made with Confidence

The second criterion is:

> Data-based decisions must be made with confidence, at least confidence in the data. This is the ultimate criterion upon which to judge the quality of data--do the decision makers who rely upon the data have confidence in them. Assuming all the lower levels of quality criteria have been met, then the final one that makes sense is that the data are actually used with confidence.

This is a good time to remind us all that confidence alone is not sufficient. One reason the construct of a hierarchy is useful is that each subsequent level depends upon earlier levels.

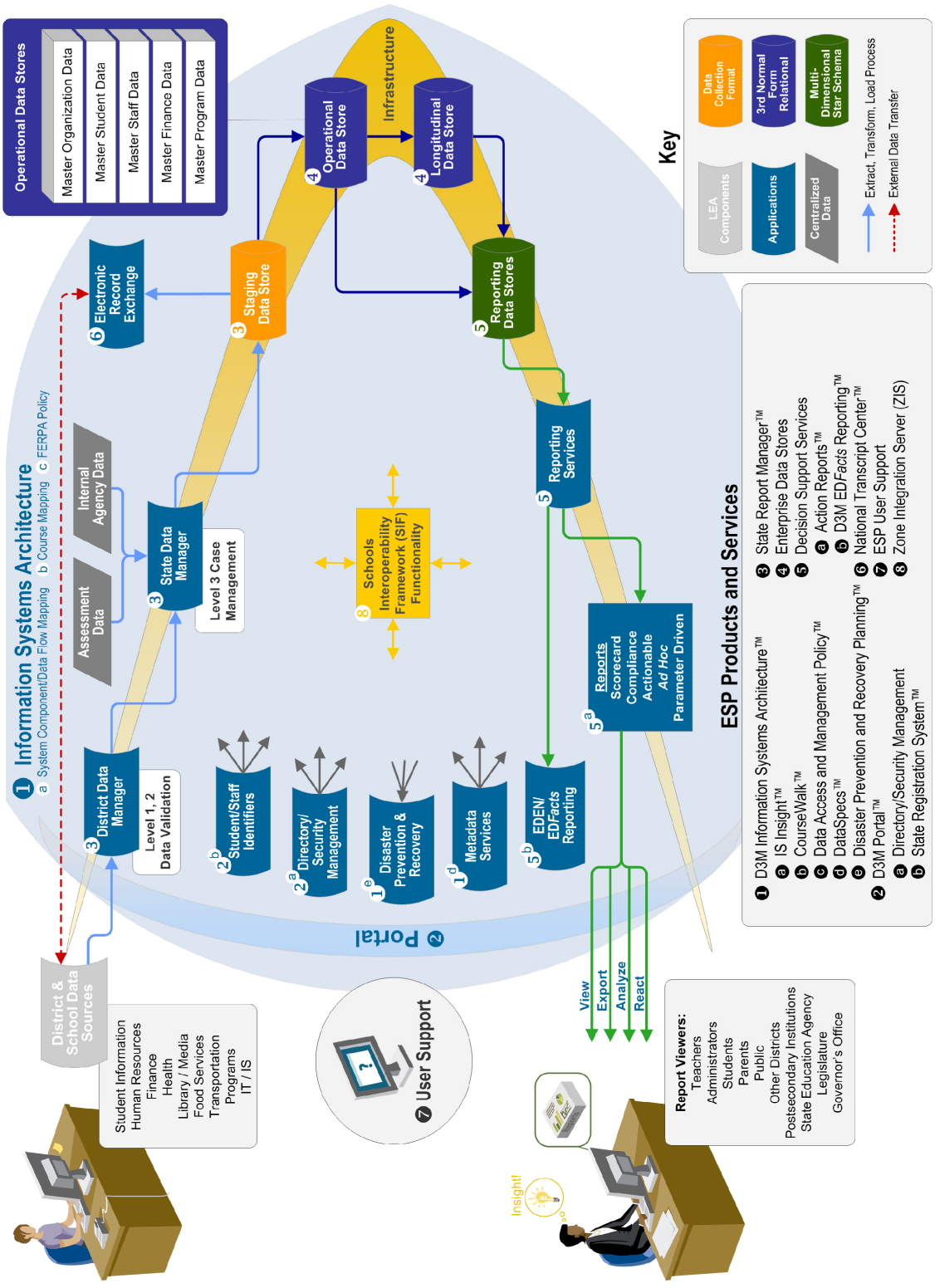
> A local district's discipline reporting system had been used for years to provide indicators of the number of students and the types of incidents in which they were involved. The reports were so clear and consistent that confidence was high. As part of a program evaluation, an evaluator went to a campus to get more details and discovered that only about 60% of all discipline incidents were routinely entered into the computer file. The others were dealt with quickly or came at a busy time. No one had ever audited a school's discipline data. On the other hand, the dropout and college-bound entries into a similar file were found to be very accurate and up-to-date.

Conclusion

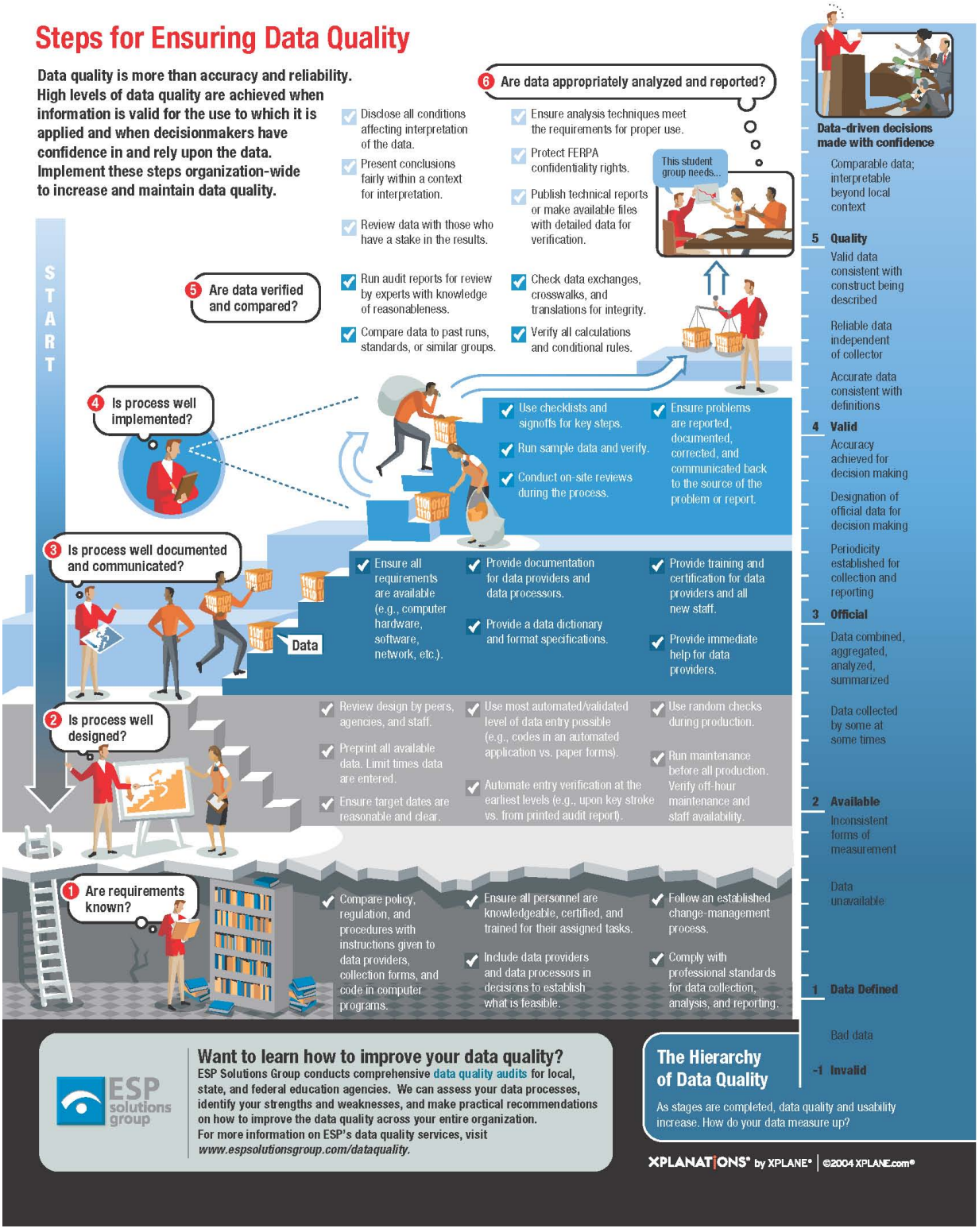
The hierarchy was a convenient way to think through what makes for quality data. Reality is that our information systems will not fall neatly into one of the levels of the hierarchy. In fact they may not often evolve sequentially through each level. At any point in time, their levels may shift up or down. What is useful here is that the hierarchy describes the characteristics of relatively low and relatively high levels of data quality.

This discussion needs to turn now to the practical side of ensuring data quality. In Part 2 of this series, The Data Quality Manual, we get down to the specific steps an education agency must follow every year to manage data for decision making.

Attachment A: ESP's D3M Framework



Attachment B: Process Illustration of Data Quality



The Optimal Reference Guide:

The Data Quality Manual, Data Quality Series – Part II

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 "dont's" 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, l, 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.

Pat	**M**	**Johnson**	**Jr**
Johnson, Pat M, Jr			

3. Be lackadaisical when the requirements are precise.

Patrick	**M.**	**Johnson**	**Jr.**
Pat		**Johnson**	

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.

Student Name	Birthday	Test Score	Course Grade	Absences
Pat Johnson	09111999	98	A	3
Pat Johnson	09111999	98	A	3
Kelly Smith	12251999	79		8
Kelly Smith	12251999		B	8

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, 2006"

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 Colleen, would you be a dear and just change those LEP codes for me again this year?"

16. Copy and paste without being extra careful.

Grade	Gender	First Name	Last Name
7	M	Freddy	Hanson
8	M	Sandra	Hernandez
7	M	Charlotte	Webster
6	M	John	Johnson
6	M	Michelle	Michelle
7	M	Juan	Paredes
7	M	Janelle	Smith
8	M	Herbert	White
8	M	Snoop	Perro

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:

- Declaration by the original source of the data (parent)
- Entry by the collector
- Compiling for reporting
- Sending
- Receiving
- Mapping
- Import
- Access
- Analysis
- Formatting
- Labeling
- Explaining
- Interpretation
- Use

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, Table 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?

Table 1: Steps for Validating Data

Step	Description	Example
1. Your Best Guess	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.)
2. Prior Statistic	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.
3. Another Entity	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.
4. Simple Math	Do some simple math with the statistic. Do the results make sense?	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.

Step	Description	Example
5. Calculate Counts	If the statistic is a percent, proportion, or ratio, calculate an actual count. Does this count make sense?	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.
6. Calculate Percents	If the statistic is a count, calculate a percent, proportion, or ratio. Does this calculation make sense?	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.
7. Know the Source	Who is reporting the statistic? Are they the right person to do so? Are they the original source? Do you trust them?	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.
8. Independent Verification	Was the statistic independently verified?	The superintendent states that 82% of the district's students passed the statewide math exam. The statistics is also reported by the state education agency and was calculated by the vendor for the assessment program.
9. Graph Proportions	If there is a graph, are the scales and proportions appropriate?	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.
10. Details and Documentation	Are definitions, measures, limitations, samples, and other information provided for judging the validity of the statistic?	The evaluation that reported the algebra enrollments is accompanied by a technical report with the details.

Step	Description	Example
11. Definitions and Periodicities	Do comparisons or changes reported use the same data points, definitions, periodicities, etc.?	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.
12. Stakes	What's at stake? How might the stakes have influenced the reporting of the statistic? How would competing perspectives have interpreted the statistic?	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.
13. Gut Reaction	What's your gut reaction?	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.

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 (Table 2). Using the criteria of validity, accuracy, lateness, usefulness, and expense, an information source can be rated on this four-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.

Table 2: Data Quality Rating Scale

Information Source: ***fill in here***				**Source Type:** ***fill in here***	
Quality Measure	**Validity**	**Accuracy**	**Lateness**	**Usefulness**	**Expense**
 High Quality	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.	Data are accurate and complete. Data standards are clear and were followed.	The most recent data are provided. The time period of the data match the use and intent of the data.	Data are presented completely and clearly for ease of use. Access to the data for use is easy.	No charge is made for access or use.
 Reduced Quality	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 nor proven.	Data standards are documented. Compliance is assumed to be reasonable. Limitations are described.	Data are recent enough to suggest reasonable applicability for use and intent.	Data are presented well for use. Access requires some effort but is available.	Copies or access is free, but some charges apply.

Information Source: *fill in here*				Source Type: *fill in here*	
Quality Measure	**Validity**	**Accuracy**	**Lateness**	**Usefulness**	**Expense**
 Low Quality	The connection between the data and the use of the data is weak or nonexistent.	Data standards are weak or nonexistent. Poor controls are in place to ensure compliance.	Aged data may not be appropriate for making decisions about current issues.	Data are poorly presented or explained. Access is cumbersome and limits use.	A charge applies for access or use.
 Poor Quality	The connection between the data and the use of the data is misrepresented or misleading.	Incorrect data, substantial missing data, or other problems are evident.	Data are too old to be useful.	Data are uninterpretable or inaccessible.	A substantial charge applies for access or use compared to similar sources of information.
 Unknown Quality	How well the data and the use of the data match is not known or not described.	Accuracy of the data is unknown or not documented.	The periodicity is unknown. The appropriateness of the data is unknown because of the lateness of them.	Unknown.	Unknown.
Information Sources: School report card, statistical report, assessment report, program evaluation, etc.			**Source Types:**		
Audiences: Public, school staff, internal agency staff, funding agency			**Perspective:** Data Provider, Data Collector, Data Reporter, Data Reseller, End User		

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

Table 3: Accuracy Scale

Accuracy		The Data are Rated at the Level in Which ALL Conditions are Satisfied.			
 High Quality	Data are accurate and complete. Data standards are clear and were followed.	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.
 Reduced Quality	Data standards are documented. Compliance is assumed to be reasonable. Limitations are described.	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.

Accuracy		The Data are Rated at the Level in Which ALL Conditions are Satisfied.			
 Low Quality	Data standards are weak or nonexistent. Poor controls are in place to ensure compliance.	41-45: A. Most key data are missing. B. Data problems are pervasive and prevent use. C. Data standards and specifications are not available.	46-50: A. Substantial, key data are missing. B. Data problems are pervasive and prevent most use. C. Data standards and specifications are not available.	51-55: 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.	56-60: 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.
 Poor Quality	Incorrect data, substantial missing data, or other problems are evident.	0-10: A. Most data are missing. B. All data exhibit major problems. C. Data standards and specifications are not available.	11-20: A. Most data are missing. B. All data exhibit problems. C. Data standards and specifications are not available.	21-30: A. Most data are missing. B. Data problems are universal. C. Data standards and specifications are not available.	31-40: A. Most data are missing. B. Data problems are substantial. C. Data standards and specifications are not available.
 Unknown Quality	Accuracy of the data is unknown or not documented.				

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 I of the Data Quality Series, *The Data Quality Imperative*, identified these four truths about data quality. They guided the design of the steps outlined below.

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 system, 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.

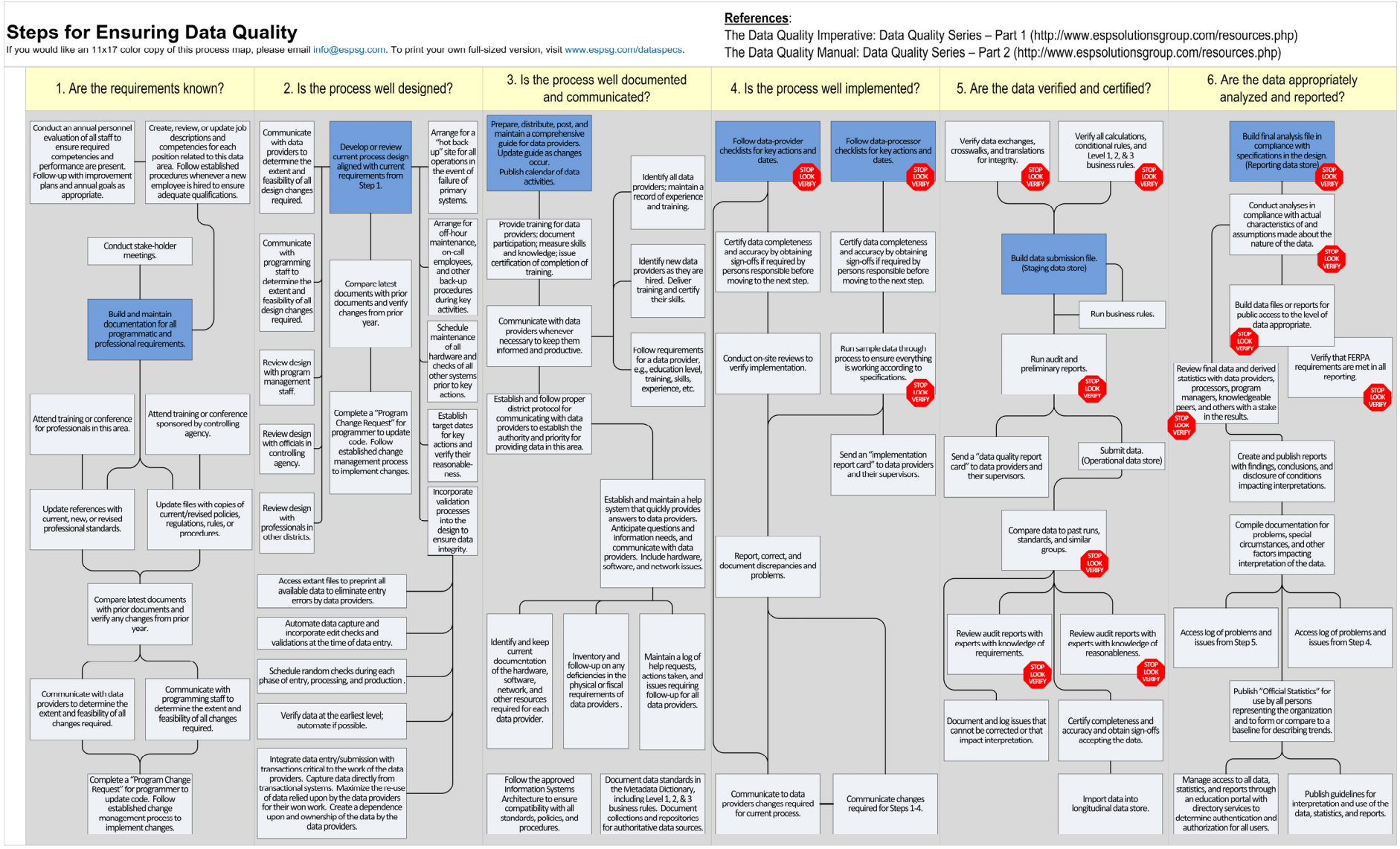
Measurement errors are those imprecisions that result from our inability to be absolutely perfect in our measurements. One is the reliability of an instrument, test, or performance task (illustrated by a test-retest difference). Measurement errors can also be "intentional" as occurs when we round numbers 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

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

Attachment A: Steps for Ensuring Data Quality

If you would like an 11x17 color copy of the process map below, please email info@espsg.com. To print your own full-sized version, visit www.espsg.com/dataspecs.



ESP Optimal Reference Guides and Optimal Reference Books

ESP covers a wide variety of education topics with our series of informational whitepapers called Optimal Reference Guides (ORGs) and Optimal Reference Books (ORBs). All are available for free download at *www.espsolutionsgroup.com/resources.php*. You can also subscribe to our monthly newsletter to have ORGs and ORBs emailed to you as soon as they are published. Just visit the link above for more information.

Data Quality

- The Data Quality Imperative, Data Quality Series—Part I
- The Data Quality Manual, Data Quality Series—Part II
- The Process for Ensuring Data Quality

Data Management

- Actions Speak Louder than Data
- From Information to Insight—The Point of Indicators
- Aligning Indicators and Actions
- Data Management Strategy for States and Districts
- Defining Data
- Management of a Education Information System
- Our Vision for D3M
- Using Assessment Results to Get Performance Results
- Why Eva Baker Doesn't Seem to Understand Accountability—The Politimetrics of Accountability

Longitudinal Data Systems

- D3M Framework for Building a Longitudinal Data System
- The Dash between PK and 20: A Roadmap for PK-20 Longitudinal Data Systems
- What's Really "In Store" for Your Data Warehouse? Data Warehouse Series—Part I
- What's Behind Your Data Warehouse, Data Warehouse Series—Part II
- Accessing Student Records in a State Longitudinal Database, Data Warehouse Series—Part III

Project Management

- Why 70% of Government IT Projects Fail, Project Management Series—Part I
- From Risk to Reward: A Guide to Risk Management, Project Management Series—Part II
- Marketing Your Field of Dreams, Project Management Series—Part III
- Project Management Success Factors

Electronic Transcripts

- Electronic Student Records and Transcripts: The SEA Imperative
- Why Your State Needs a PK-20 Electronic Record/Transcript System

Standards

- Articulating the Case for Course Numbers
- Confidentiality and Reliability Rules for Reporting Education Data
- FERPA: Catch 1 through 22
- Graduation Rates: Failing Schools or Failing Formulas?
- National Education Data Standardization Efforts
- Racial/Ethnic Data Reporting in Education
- Recommended Data Elements for EDEN Reporting
- Revisions to FERPA Guidance

Trends in Education

- Data-Driven Decision Making 2016
- How Education Information Fared in the Last Decade
- IT Defined…for the Educator
- Why My Space Matters to the K-12 Space

Student/Staff Identifiers

- Requirements for an RFP for Student Identifiers
- Statewide Student Identifier Systems

Disaster Prevention & Recovery

- Disaster Prevention and Recovery for School System Technology

Growth Models

- Growth Model Growing Pains, Growth Model Series—Part I
- Comparison of Growth and Value-Add Models, Growth Model Series—Part II
- Making a Year's Growth and Performing on Grade Level: Muddled Definitions and Expectations, Growth Model Series—Part III
- Growth Models—Finding Real Gains

ESP Solutions Group

(512) 879-5300
www.espsolutionsgroup.com