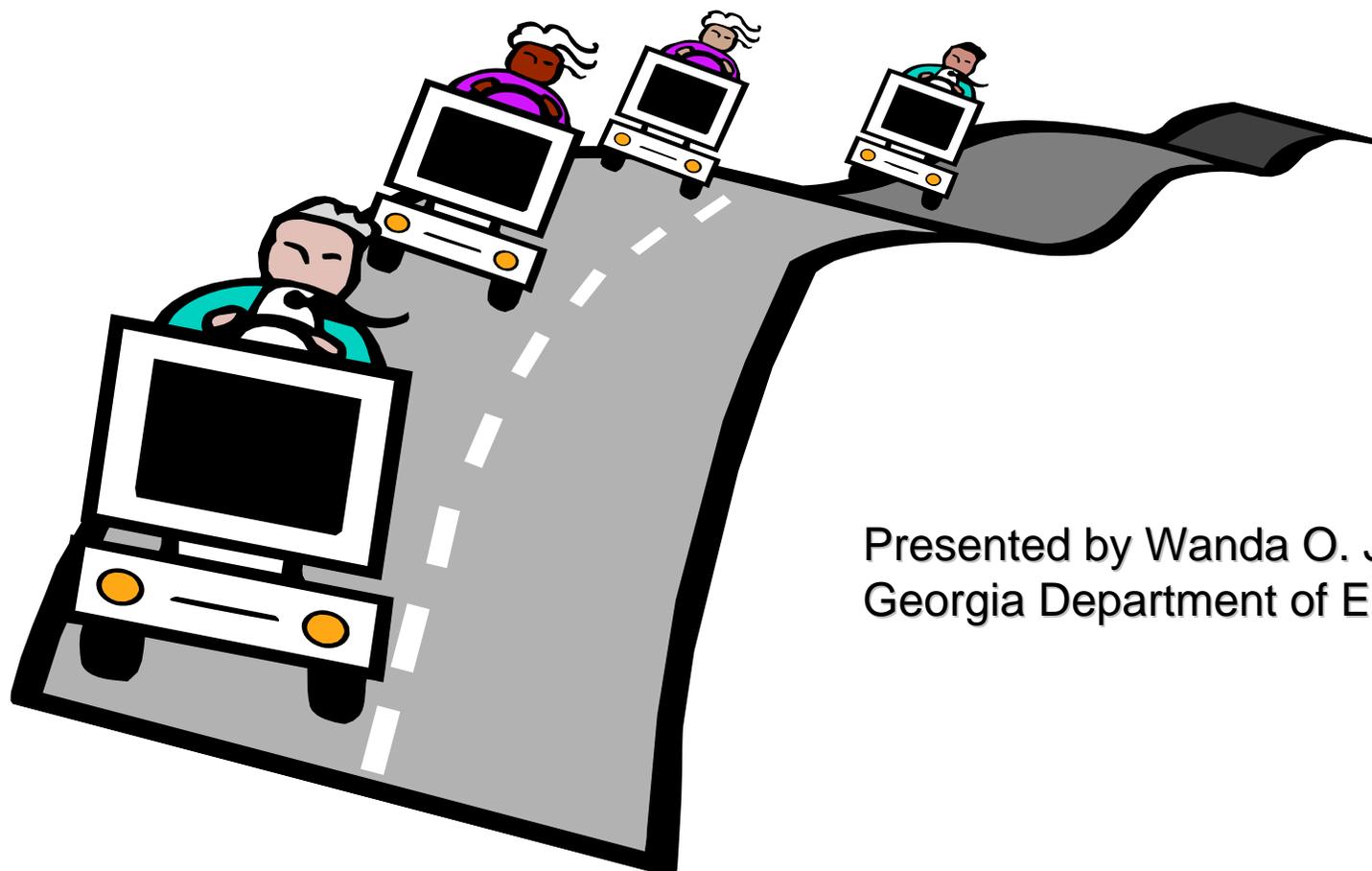


Destination: Quality Data A Process, Not A Place



Presented by Wanda O. Jones
Georgia Department of Education

Data Collections & Reporting Team

Levette Williams, Director

Beverly Ellison, Secretary

Data Collections

- ▶ Wanda Jones, Lead Specialist
- ▶ Jason Whitley, Analyst
- ▶ Betty Rickikki, Specialist
- ▶ Kathy Aspy, GTID Administrator

Data Reporting

- ▶ Nick To, Specialist
- ▶ Mark Vignati, Analyst

Data Quality Overview

Organizations and agencies often find it difficult to rely on the information provided through internal data collection and reporting processes.

Yet, it is this information that serves as the foundation for internal decision-making.



Data Quality Overview



What is data quality?

Why is it important?

What is the impact of bad data?

What does this mean to Georgia?

What has Georgia done to improve data quality?

Definition of Data Quality

The level of excellence of data when compared to certain criteria. Criteria may vary among organizations.

Common factors or criteria include:

- ▶ Data Consistency
- ▶ Data Accuracy
- ▶ Data Integrity
- ▶ Business Rules
- ▶ Completeness
- ▶ Redundancy
- ▶ Meaningful to Users
- ▶ Availability
- ▶ Metadata Documented
- ▶ Relevancy

Why Data Quality?

Data quality is important to an organization for several reasons:

- ▶ Credibility of the data
- ▶ Credibility of the organization
- ▶ Impact on internal projects, policies, procedures, and funding
- ▶ Less time spent on rework



Why Data Quality?

- ▶ Lower costs to organization

A 2002 study performed by the Data Warehouse Institute (TDWI) estimates that poor data quality costs businesses in the United States an excess of \$600 billion each year.

That same study indicated that data deteriorates or becomes obsolete at a rate of 2% every month.

Dyche', Jill. "The Bottom Line on Bad Customer Data." Baseline Consulting Group, Inc. (2004) Retrieved on February 20, 2007 from the Data Warehouse Institute website: <http://www.tdwi.org/Marketplace/WPContent.aspx?PID=170>



Impact of Bad Data

Example #1

In 1974, an 18 year old man tried to take what he thought was scrap metal from the parking lot of a machine shop. He ended up spending part of his summer vacation in jail for misdemeanor theft.

Thirty years later, a background check of his criminal record sold to employers by a leading identification and credential verification service erroneously reported his misdemeanor as a felony. The report also indicated that he spent seven years in prison when he spent 51 days in county jail.

This data error cost him several job opportunities.

Impact of Bad Data

Example #2

“In an apparent and rare in-house critique, the president of the Federal Reserve Bank of Dallas said that because of faulty inflation data, the Fed kept interest rates too low for too long earlier this decade, fueling speculative housing activity.”

Ip, Greg. “Official Says Bad Data Fueled Rate Cuts, Housing Speculation.” Wall Street Journal Online. Retrieved from the Real Estate Journal.com on February 16, 2007:
<http://www.realestatejournal.com/buysell/mortgages/20061106-ip.html>

Impact of Bad Data

Example #3

In England, the criminal record of a man suspected of making sexual advances to 11 under age girls went undetected by the local police. The man was subsequently employed as a school caretaker at a local school. In 2003, he was convicted of murdering two girls.

After his conviction, authorities learned that the details of his criminal record passed through 4 police systems and the Police National Computer.

Infoshare. (2007). "Why Data Quality is So Important".. Retrieved February 19, 2007, from the Infoshare Limited website: http://www.infoshare.ltd.uk/company/why_data_quality_is_so_important.html

Universal Impact

The previous examples look at different industries and describe the impact of bad data on the lives of various individuals.

Are data quality issues isolated to certain industries, geographic locations, or population groups?

The answer is a resounding **NO!**

All organizations, regardless of industry, are vulnerable to challenges with data.

Education systems are no different. Data quality issues occur everywhere.

Data Quality in Education

What kind of data quality issues have we seen in Georgia's education systems?

- ▶ Under reporting of data
- ▶ Over reporting of data
- ▶ Occurrences of invalid data
- ▶ Occurrences of inaccurate data
- ▶ Incomplete data



Under Reporting of Data

Example

A school district failed to report 16 principals during a data collection that generated funding based on training and experience.

This data reporting error resulted in a loss of funding of more than \$736,000 in base pay alone. Thousands of dollars in additional funding was loss after accounting for the different levels of certification and creditable years of service for each person.

Under Reporting of Data

Example

During a comparison of federal and state data, it was noted that some school districts failed to report hundreds of students that had enrolled due to hurricane Katrina. They had reported higher numbers in the federal report than in the state report.

Over Reporting of Data

Example

ABC High School reported 30 discipline incidents for non-felony drugs involving 10 students at different times throughout the school year. When the data was questioned, it was discovered that there were only 10 incidents. The school had misinterpreted the definitions and rules for reporting incidents.

Invalid Data

Example

Students in third grade reported as dropouts due to incarceration.

Students reported as a male in one data collection cycle and later reported as female.

Certified teachers reported with a birthdate that would make them 8 years old.

Inaccurate Data

Example

Jane Doe was reported as an active student at ABC elementary school in the fall enrollment count used for state funding. During the accountability collection at the end of the school year, Jane Doe was not reported by ABC elementary at all.

ABC school received an error message because they are required to account for all students reported in the fall. ABC commented that they had reported Jane Doe in error. She had enrolled and was expected to attend, but never actually showed up at school.

Impact: Jane was included in enrollment counts and funding, but did not really attend that school.

Incomplete Data

Example

School ABC reported only 3 out of 6 instructional segments for every 3rd grade student (approximately 450 students). Instructional segments are used to determine funding.

This resulted in the lost of approximately \$1270 per student plus additional funds allocated to various instructional programs (minimum loss of \$571,500).

Resolution of data issues

Given the need for quality data, and the impact of data on decision-making processes, what has Georgia done to resolve data issues?

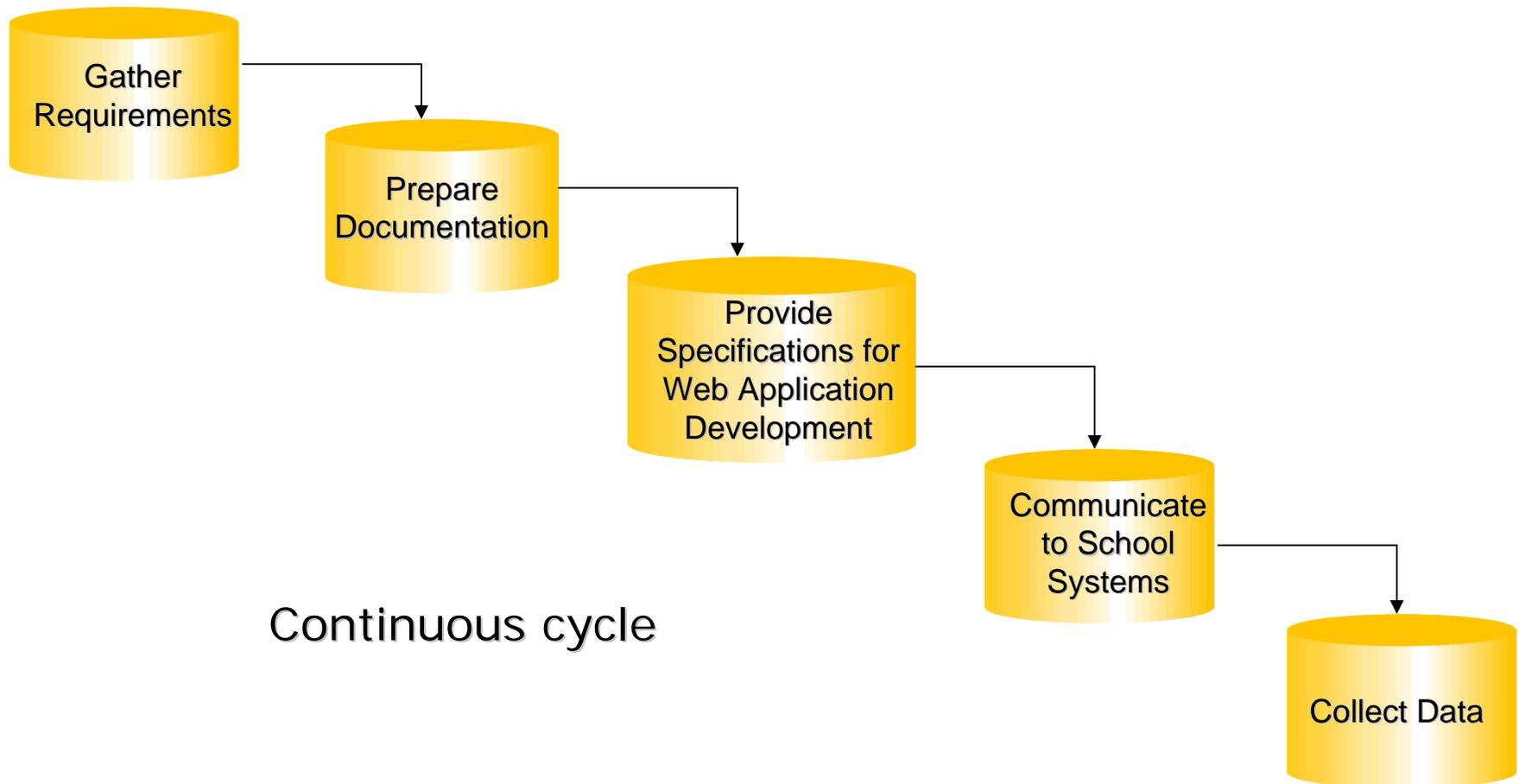


Data Collection Process



Five Steps for Georgia's Data Collection Process

Data Collection Process



Data Collection Process

► Requirements driven by:

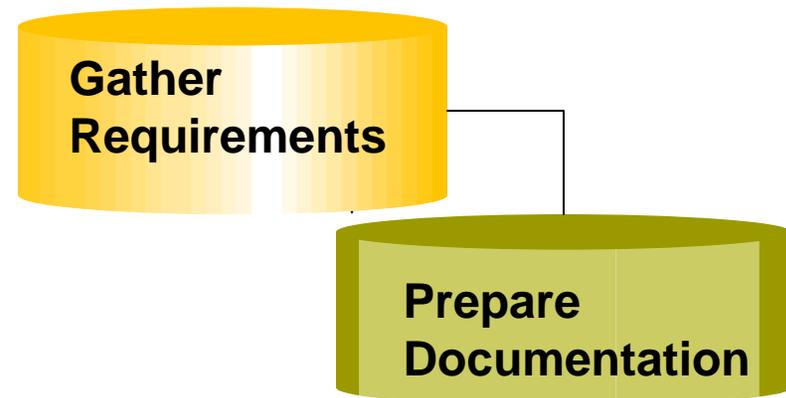
- Changes in federal requirements
- Changes in state laws
- Changes in board rules
- Data needed for program evaluation, enforcement, or audit



Data Collection Process

► Types of documentation:

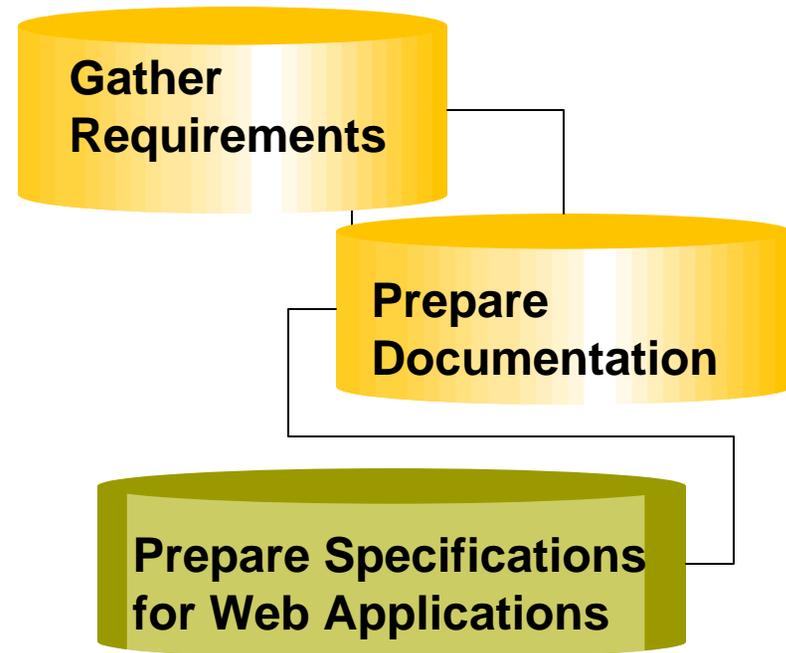
- Business Rules
- File Layouts
- Data Element Definitions
- Transmission Dates
- Transmission Instructions
- Coordinator's Guide
- Use Cases
- Vendor Notes and Data Transformations



Data Collection Process

► Specifications for web applications:

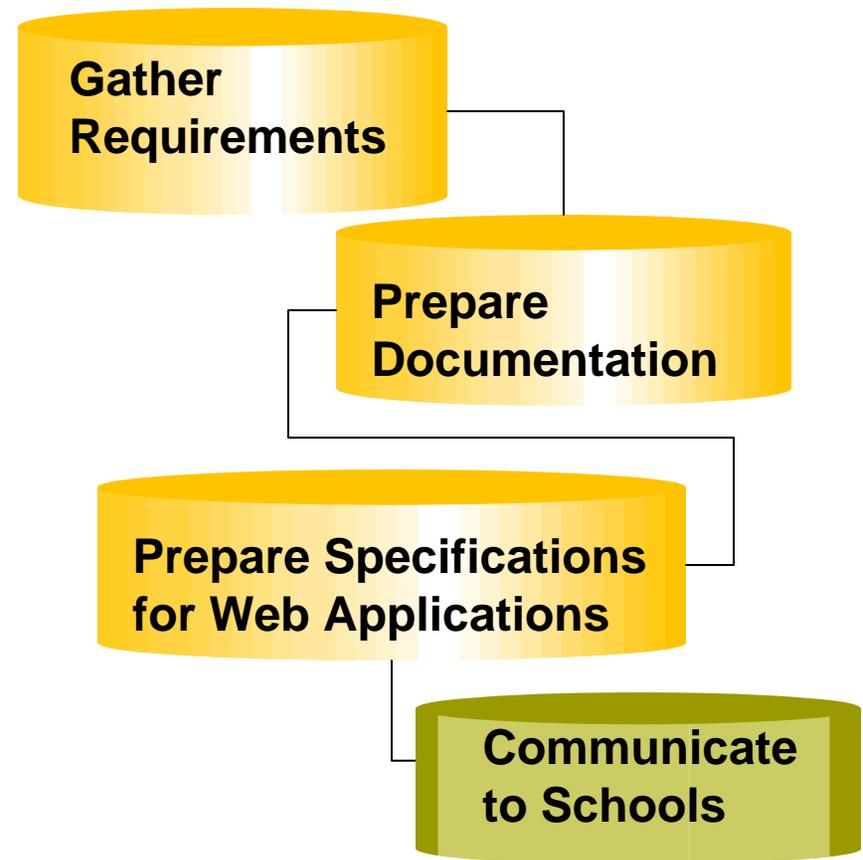
- Cross-functional team
 - Web Development
 - Quality Assurance
 - Database Administration
 - Infrastructure
- User Acceptance Testing



Data Collections Process

► Training of Coordinators in 184 School Districts (2,159 schools)

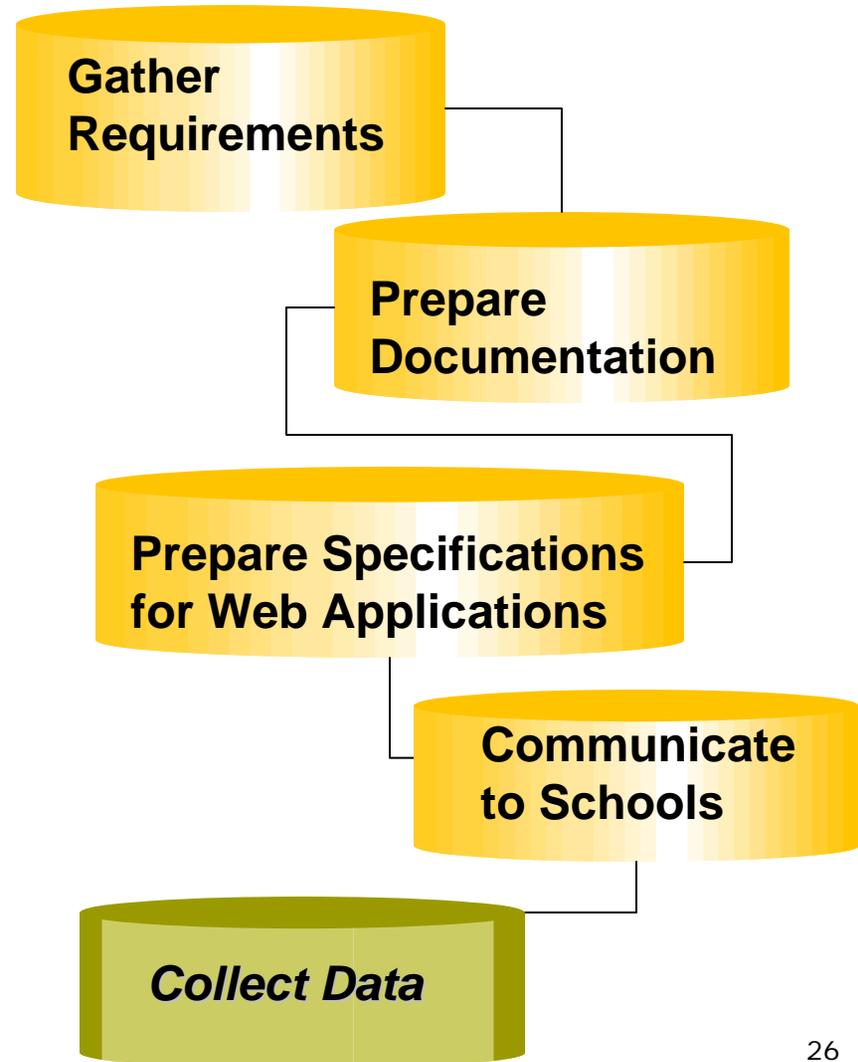
- Education Technology Training Centers
- 2007 – 8 two-day workshops with over 875 participants
- Partnered with program managers from other areas



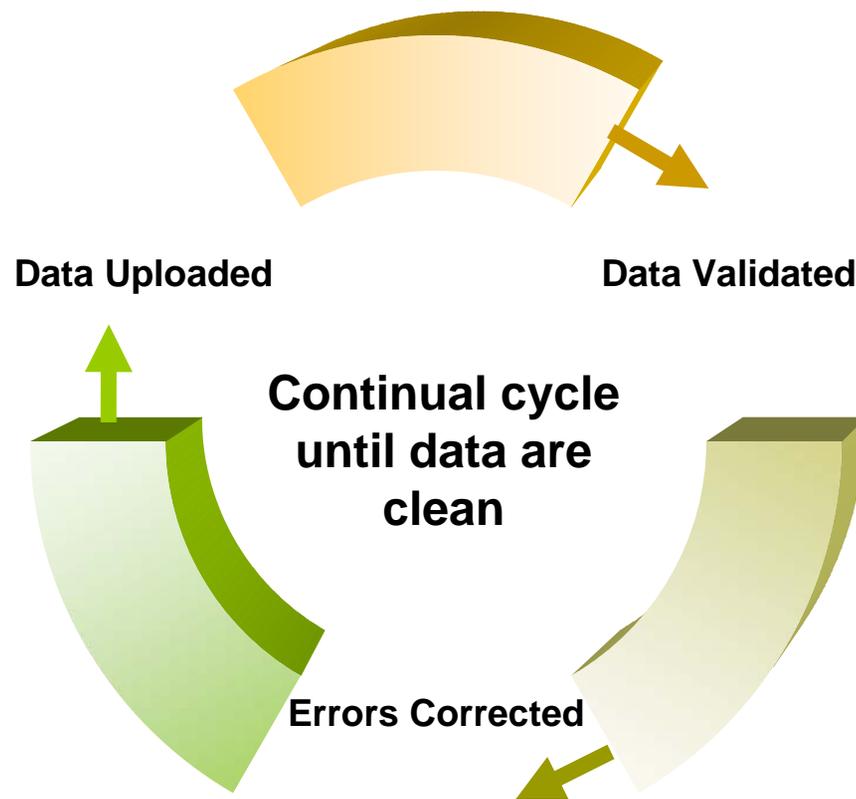
Data Collections Process

► Data Collection:

- Daily monitoring of data
- Error Reports and Verification Reports for school districts and state agency
- Help desk staff as first level support for school districts



Data Collection Process



Data Collection Process

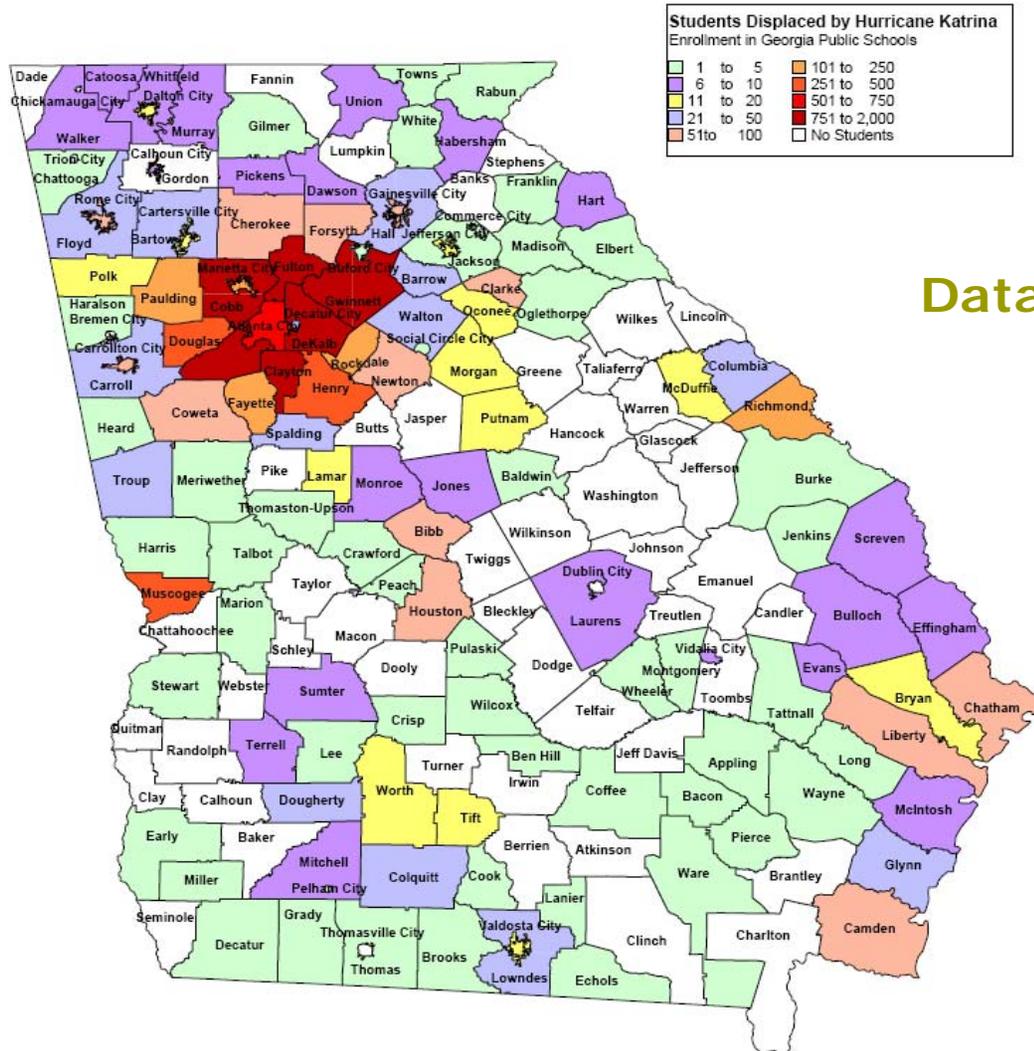
- ▶ School systems have final reports to review and copy (if desired)
- ▶ School system superintendents must electronically sign-off the data certifying that the data are accurate
- ▶ Post Data Collection procedure
 - Data scrubbing (Quality Control)
 - Data are securely stored in database tables

Data Collections- Records Collected

Data Collection	# of Records Collected
FTE (cycles 1 and 3)	~ 3.3 million records
FTE (cycle 2)	~ 250,000 records
CPI	~ 1.4 million records
Class Size	~ 264,200 records
Student Record	~ 15 million records



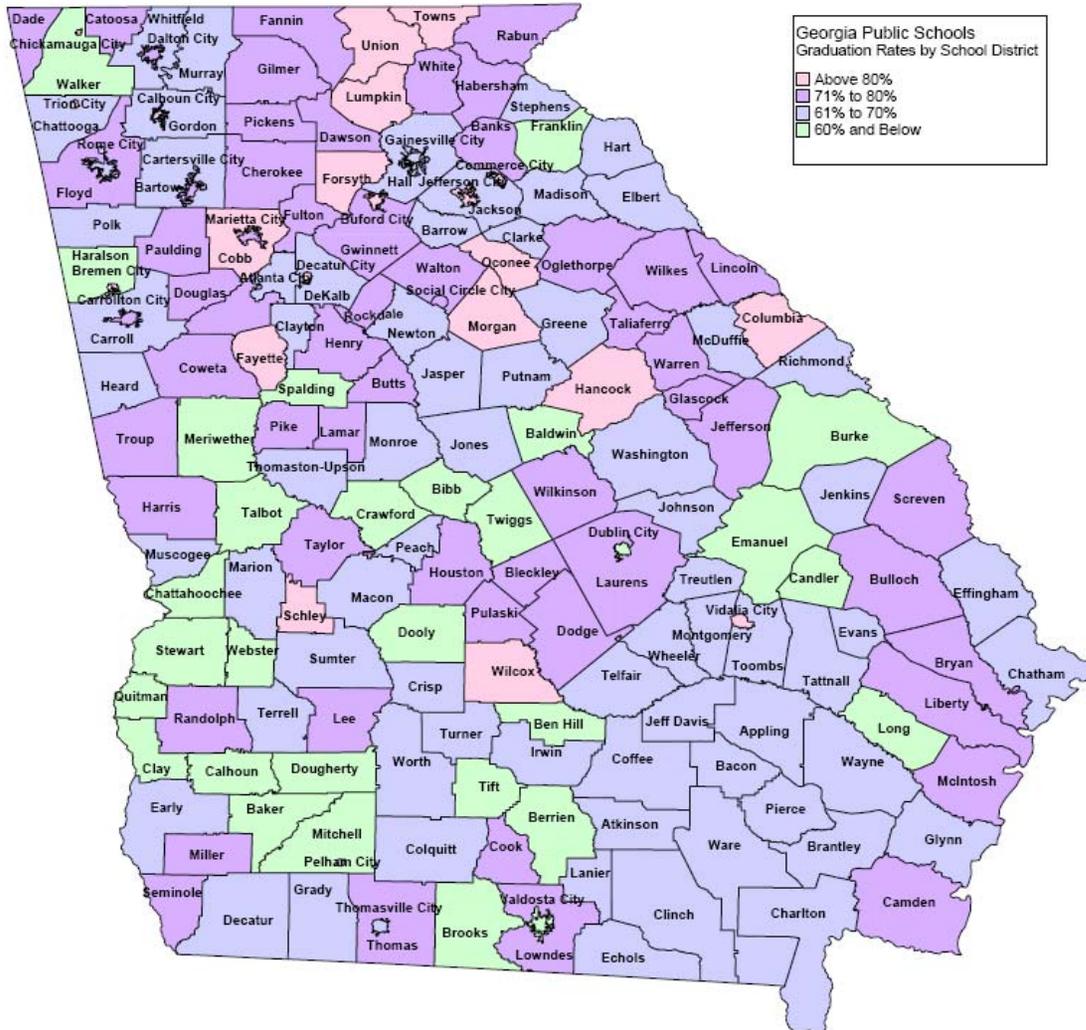
**STUDENTS DISPLACED DUE TO HURRICANE KATRINA
ENROLLMENT IN GEORGIA'S PUBLIC SCHOOLS
OCTOBER 2005 - TOTAL STUDENTS: 10,290**



Data Reporting



Graduation Rates By School District School Year 2005-2006



Data Reporting

Continuous Improvement

- ▶ Scrub Data
- ▶ Load Data Warehouse
- ▶ Identify More Effective Processes
- ▶ Utilize technology advancements
- ▶ Encourage Feedback from Customers

Points to Remember

- ▶ Recognize that data quality issues exist in every organization
- ▶ Be willing to acknowledge data issues when identified and establish a process to address data issues
- ▶ Conduct periodic assessments of data quality and take action accordingly

Consider This ...

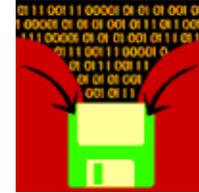
- ▶ According to Forrester research, global IT spending will reach \$2.1 trillion in 2007.
- ▶ Although this represents a 6% increase from 2.02 trillion in 2006, it is a slowdown in the rate of growth from 8% in 2006.

Bartel, Andrew. "Global IT Spending and Investment Forecast, 2006 to 2007." Forrester Research, Inc. Retrieved from Forrester.com website on February 27, 2007:
<http://www.forrester.com/Research/Document/Excerpt/0,7211,40451,00.html>

Consider This ...

How many companies/organizations have a line item in the budget for data quality or related costs?

Moral of Story ...



Data is to your reporting system as blood is to your body.

No blood → Dead body (shell)

No data → No information (shell)

Contaminated blood → Sick body (bad output)

Contaminated data → Data issues (bad output)

Questions & Answers

(see Levette for hard questions!)

