One-Voice for Sewer Condition Assessment Data

Dr. Phil Lewis, Dr. Yongwei Shan, Hossein Khaleghian

Oklahoma State University

Introduction

- This research is a collaborative project between the National Association of Sewer Service Companies (NASSCO) and Oklahoma State University (OSU).

- The goal is to develop a unified national inventory of underground sewer condition assessment data.

- The outcomes will allow sewer stakeholders to have access to existing sewer condition data from across the nation to benchmark sewer infrastructure performance.

Background

- Quality data of current condition of sewer pipelines is fundamental for the development of sewer asset management tools and strategies.

- Factors such as age, material, and soil type that have significant effects on pipe condition were determined.

- However, quality issues in sewer inspection data and condition ratings were not addressed.
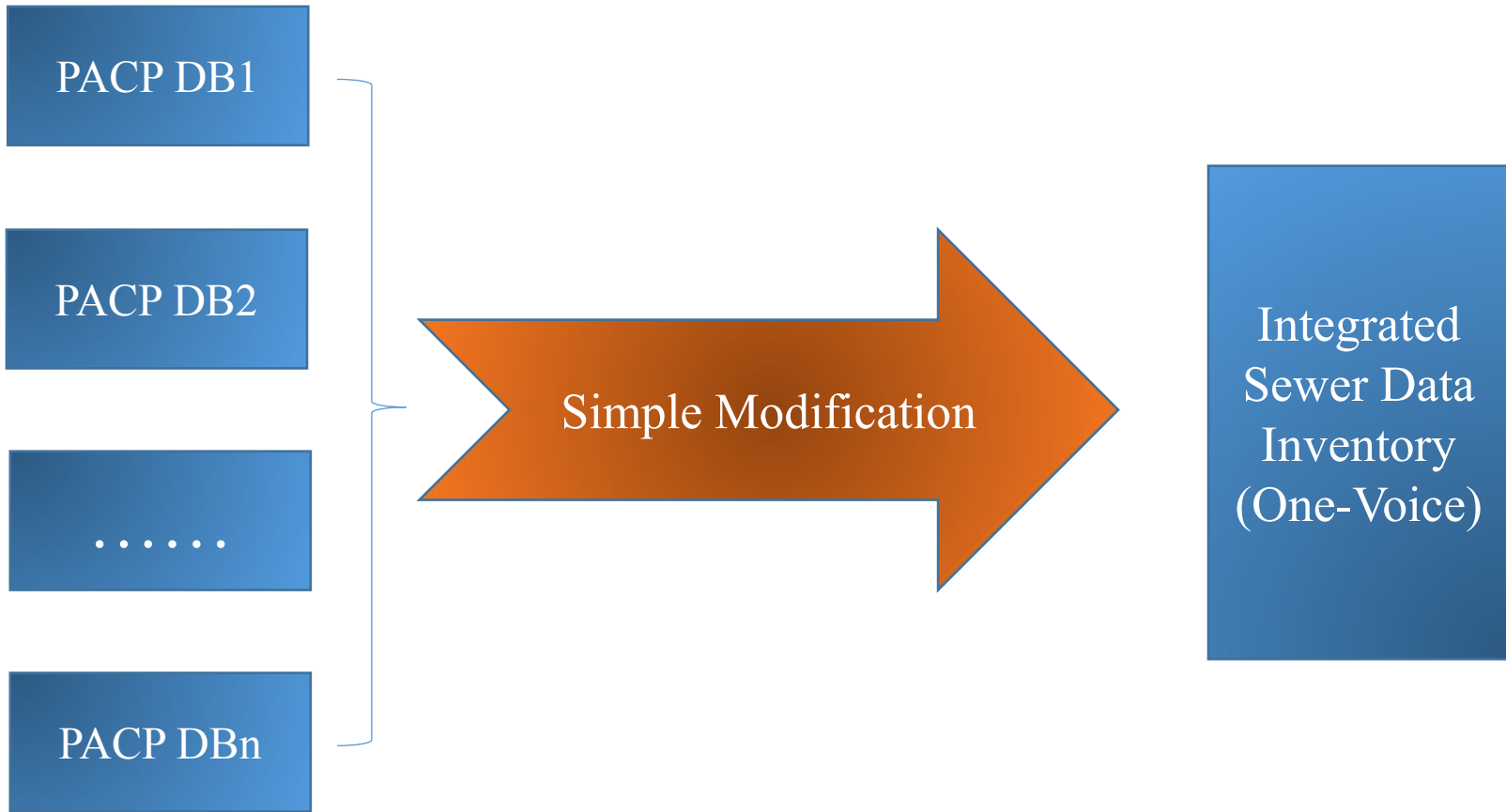
PACP

- In 2002, NASSCO developed the Pipeline Assessment and Certification Program (PACP) as a standard data collection format for Closed-Circuit Television (CCTV) inspections.

- PACP became the industry standard for sewer condition data and it was implemented by more than 200 cities and utility districts.

- The objective of PACP is to evaluate the internal structural and operational condition of sewer pipelines.

- PACP provides a grading system to quantify pipe conditions based on the most severe defects (Quick Rating) or the average severity of grades (Pipe Rating Index).
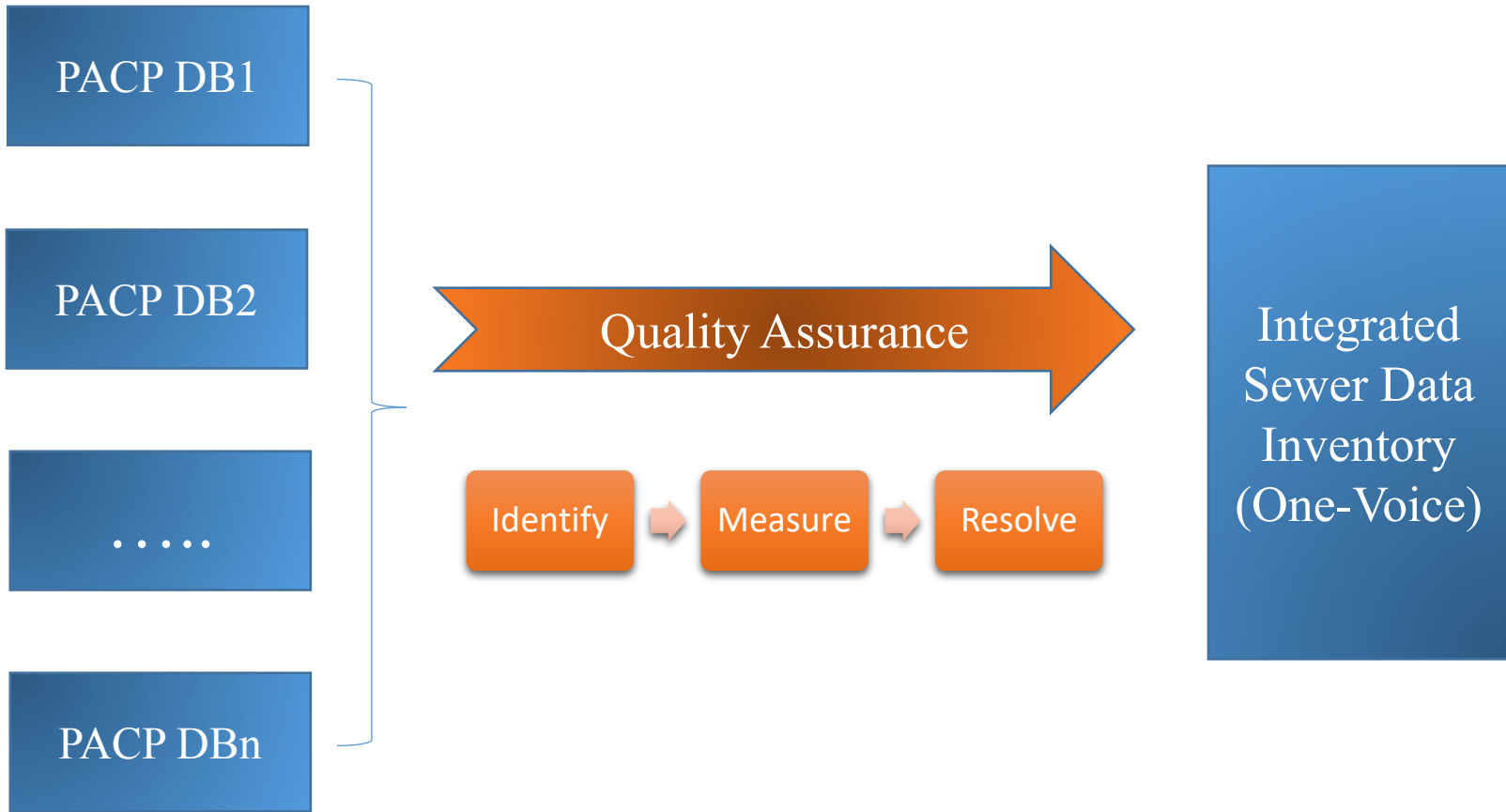
Research Approach

- Collecting PACP Data

- Identifying data quality issues in PACP databases.

- Developing data quality assurance process to address those issues:

  ➢ Data Quality Metrics

  ➢ Data Quality Framework

  ➢ PACP Grading System

Original Project Objectives

PACP DB1

PACP DB2

. . . . . .

PACP DBn

Simple Modification

Integrated Sewer Data Inventory (One-Voice)

Updated Project Objectives

PACP DB1

PACP DB2

. . . . .

PACP DBn

Quality Assurance

Identify → Measure → Resolve
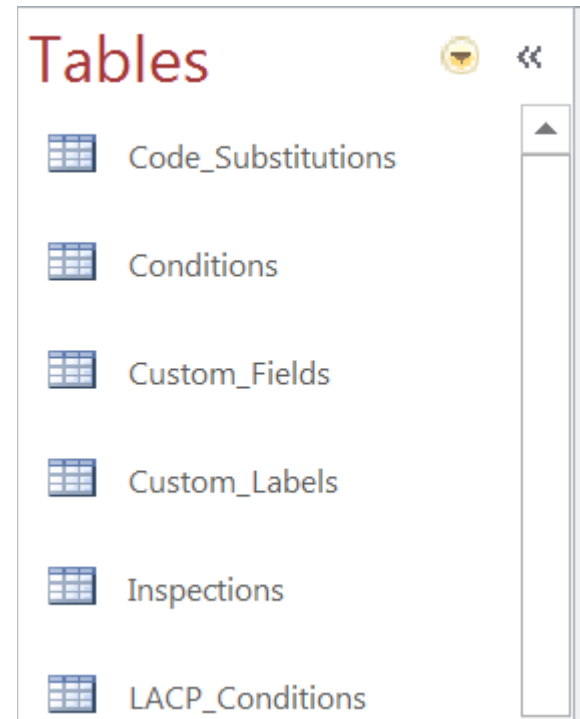
Integrated Sewer Data Inventory (One-Voice)

Data Collection

- A confidentiality agreement was drafted to ensure anonymity to data providers.

- Data transfer protocols were developed to ensure security of the data storage.

- To date, six datasets have been collected and two different PACP data collection software programs were evaluated.

- Based on six datasets, the research team identified differences in data management practices among the data providers.

First Database (FW)

- Coding System: PACP 6
- Data Structure: PACP 6
- Inspections: 5232
- Conditions: 84785

- Main Problems:
  - Duplications

Tables

- Code_Substitutions
- Conditions
- Custom_Fields
- Custom_Labels
- Inspections
- LACP_Conditions

Second Database (SUD)

- Coding System: PACP 6

- Data Structure: Software Preference

- Inspections: 212

- Conditions: 1916

- Main Problems:
  - Duplications
  - Data Structure
  - 121 Input Errors

Tables

- Info
- L
- LI
- LO
- Media
- Media_Path
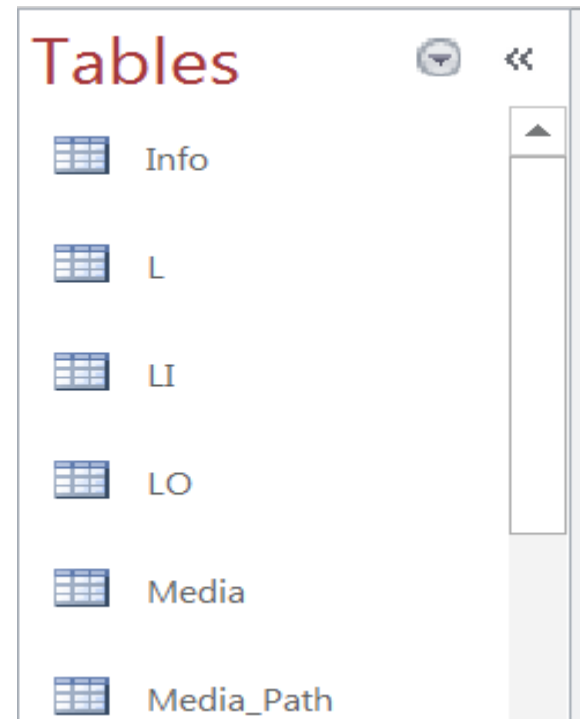
Third Database (LC)

- Coding System: PACP 6
- Data Structure: PACP 6
- Inspections: 2996
- Conditions: 28405

- Main Problems:
  - o Duplications
  - o Separate Databases

Large Diameter Interceptors 2015.MDB
LASA 2008.MDB
LASA 2009.MDB
LASA 2010.MDB
LASA 2011.MDB
LASA 2012.MDB
LASA 2013.MDB
LASA 2014.MDB
LASA 2015.MDB
Mid Size Interceptors 2015.MDB

Forth Database (TU)

- Coding System: PACP 4.2
- Data Structure: PACP 4.2
- Inspections: 46091
- Conditions: 365659

- Main Problems:
  - o Incompatibility

Tables

- Code_Substitutions
- Conditions
- Custom_Fields
- Custom_Labels
- Inspections
- Media_Conditions

fdImpExp

Unexpected error in Module: Import_Export , Function: ValidatePACPv42

3265, Item cannot be found in the collection corresponding to the requested name or ordinal.

Contact the Help Desk if you require further information

OK

Fifth Database (LA)

- Coding System: PACP 6
- Data Structure: Software Preference
- Inspections: 7587
- Conditions: 99596

- Main Problems:
  - Separate Databases
  - Data Structure

CCTV_2015PACP.mdb
CCTV_2016PACP.mdb

Tables

AI_T   Table

Date Created: 7/29/2008 3:36:27 PM
Date Modified: 4/15/2016 9:50:32 AM

AIC_T   Table

Date Created: 7/29/2008 3:36:27 PM
Date Modified: 4/15/2016 9:50:32 AM

C_T   Table

Date Created: 7/29/2008 3:36:27 PM
Date Modified: 8/13/2010 7:53:24 AM

DC_T   Table

Date Created: 7/29/2008 3:36:27 PM
Date Modified: 7/29/2008 3:36:28 PM

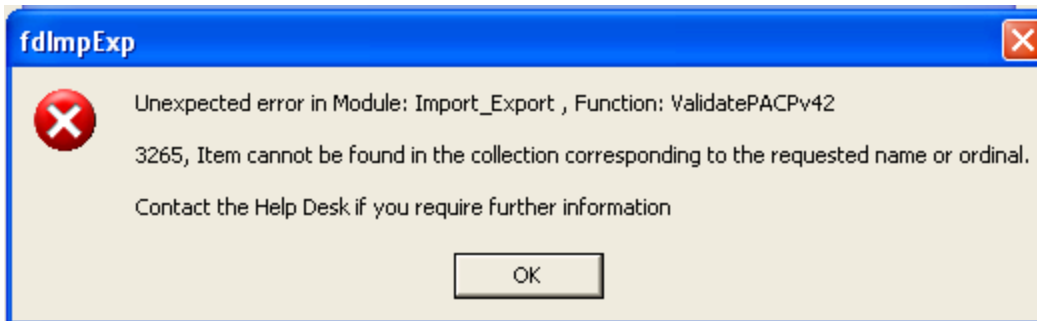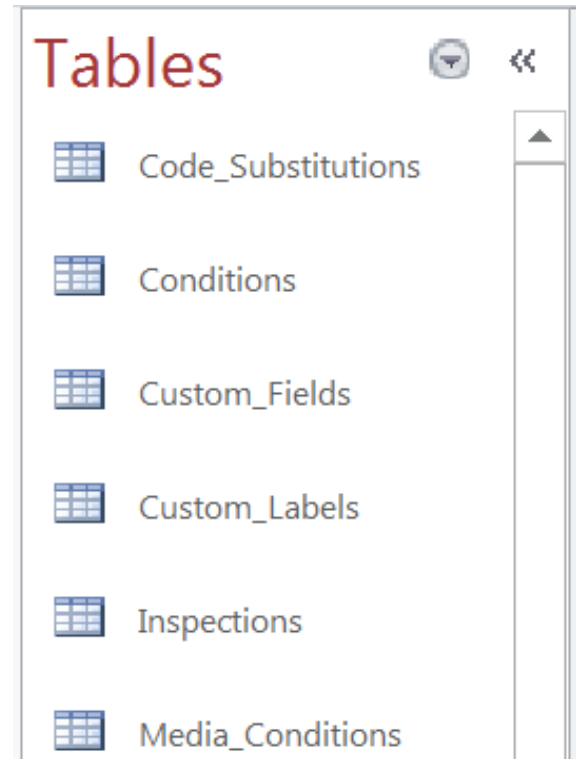# Sixth Database (BA)

- Coding System: PACP 6

- Data Structure: PACP 6

- Inspections: 40966

- Conditions: 522400

- Main Problems:
  - 22,084 Input Errors /12,115 Inspections
  - Separate Databases

Tables

- Code_Substitutions
- Conditions
- Custom_Fields
- Custom_Labels
- Inspections
- LACP_Conditions

StandardPacpExchange.mdb
StandardPacpExchange2.mdb
StandardPacpExchange3.mdb
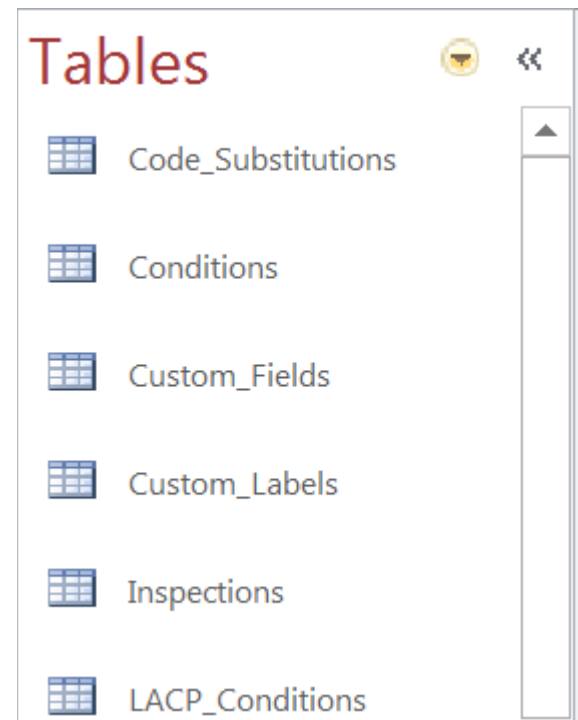
Data Quality Assurance

- The goal of data quality assurance is to reach a high level of accuracy in the PACP inspection data and make it consistent with other datasets.

- This process is a significant step in developing a sewer data inventory by integrating existing datasets.

- High quality data are necessary for decision-support systems, design analysis, and research.

- Technology-only approaches are not sufficient to provide sustained data quality improvements.

- Data quality cannot be improved independently of the source or the context in which these data are used.

Data Quality Assurance

- In order to develop proper QA procedures for PACP datasets, the collected data were reviewed to detect any data quality problems.

- By reviewing the PACP data with the help of NASSCO consultants, the major issues in the database were determined.

  - ➤ Pipe grading system
  - ➤ Duplicated data

Data Quality Process

Identify

- Database Rules (PACP, …)
- Develop Data Quality Flowchart
- Define Data Quality Metrics

Measure

- Assess Data
- Interpret Results
- Develop Reports and Scorecards

Resolve

- Identify Root Cause
- Define Correction Process
- Implement Resolution
  - ➢ Do Nothing
  - ➢ Correct
  - ➢ Replace
  - ➢ Combine
  - ➢ Remove

17
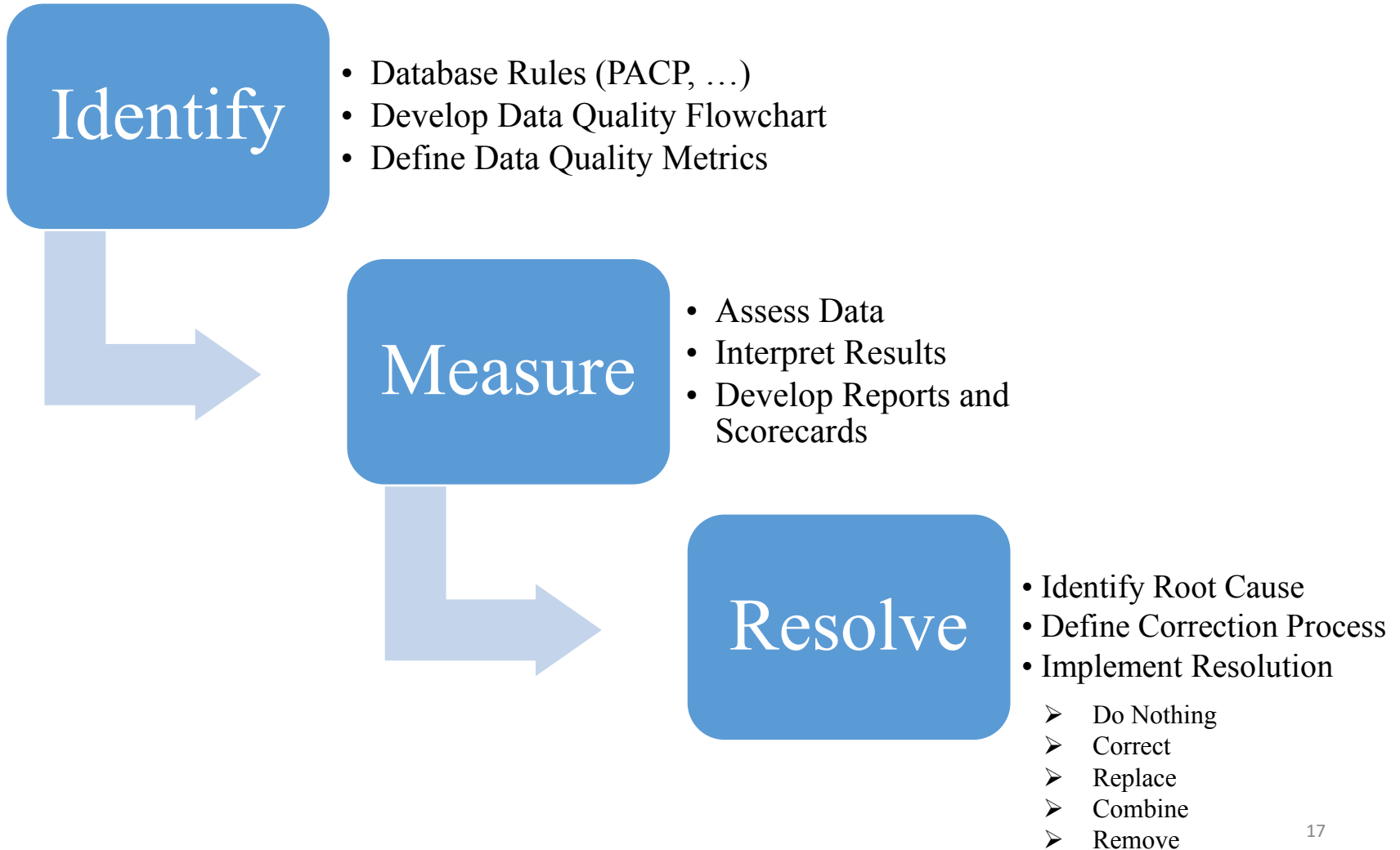
Sewer Inspection Data Quality Concerns

- Data developed by different inspectors and not consistent

- Software incompatibility

- Increase use of data as a decision support tool

- Accessibility of data

- Implementation of the PACP standards

Data Quality Metrics

- Metrics should be insensitive to changes in the number of records in the database;

- Metrics should accurately reflect the degree to which the data meets the associated data quality need;

- Metrics should be independent of each other, so that no two metrics are actually measuring the same effect; and

- The number of metrics chosen should be kept to a reasonable number, as too many metrics can often confuse rather than clarify.

- Metrics should address PACP database rules!

Name	Description
Validity	Data element passes all edits for acceptability
Completeness	Data element is required based on the condition of another data element and database rules
Consistency	Data element is free from variation and contradiction based on the condition of another data element
Uniqueness	Data element is unique (duplicate values)
Timeliness	Data element represents the most current information
Accuracy	Data element values are properly assigned

PACP Rules

Code	Value S/M/L	Value 1st	Value 2nd	Value %	Clock At/From	Clock To
CC	NR	NR	NR	NR	R	R
CL	NR	Length (O)	NR	NR	R	NR
CM	NR	NR	NR	NR	R	R
CH2	NR	NR	NR	NR	R	R
CH3	NR	NR	NR	NR	R	R
CH4	NR	NR	NR	NR	R	R
CS	NR	NR	NR	NR	R	R
FC	NR	NR	NR	NR	R	R
FL	NR	Length (O)	NR	NR	R	NR
FM	NR	NR	NR	NR	R	R
FH2	NR	NR	NR	NR	R	R
FH3	NR	NR	NR	NR	R	R
FH4	NR	NR	NR	NR	R	R
FS	NR	NR	NR	NR	R	R
B	NR	O	NR	NR	R	O
BSV	NR	O	NR	NR	R	O
BVV	NR	O	NR	NR	R	O
H	NR	O	NR	NR	R	O
HSV	NR	O	NR	NR	R	O
HVV	NR	O	NR	NR	R	O
D	NR	NR	NR	R	NR	NR
DH	NR	NR	NR	R	NR	NR
DV	NR	NR	NR	R	NR	NR
XP	NR	NR	NR	R (>=40%)	NR	NR

PACP Rules (Example)

B	193
-Should not have a value in Value_Percent	193
BSV	19
-Should not have a value in Value_Percent	19
BVV	53
-Should not have a value in Value_Percent	53
CC	299
-Should not have a value in Value_Percent	299
CH2	7
-Required field Clock_To missing	4
-Should not have a value in Value_Percent	3
CL	474
-Should not have a value in Value_Percent	474
CM	543
-Should not have a value in Value_Percent	543

Duplicates

- Common practices to resolve duplicates:

1. Eliminate duplicate entries (inspection records for pipes with same inspection date (or age) and structural condition rating).

2. Eliminate inconsistent inspection records (inspection records for pipes with same inspection date (or age) but different condition ratings).

Salman, B. (2010). *Infrastructure management and deterioration risk assessment of wastewater collection systems*

(Doctoral dissertation, University of Cincinnati)

Duplicates

	Inspections														
InspectionID	Pipe Segment	Date	Time	Street	USMH	Up_Rim_to_Invert	DSMH	Dn_Rim_to_Invert	Direction	Height	Shape	Material	Total_Length	Length_Surveyed	Location_Code
284	C10 007 Y	20070321	12:34	Lincoln Pkwy	C10 007	10.5	C10 008	0	Downstream	10	Circular	Vitrified Clay Pipe	242	242	Light Hway
3429	C10 007 X	20120319	12:50	6935 Lincoln Pkwy	C10 007	11.5	C10 008	12	Downstream	10	Circular	Vitrified Clay Pipe	?	86.7	Yard
3431	C10 007 X	20120320	8:38	6935 Lincoln Pkwy	C10 007	11.5	C10 008	12	Upstream	10	Circular	Vitrified Clay Pipe	?	154.5	Yard
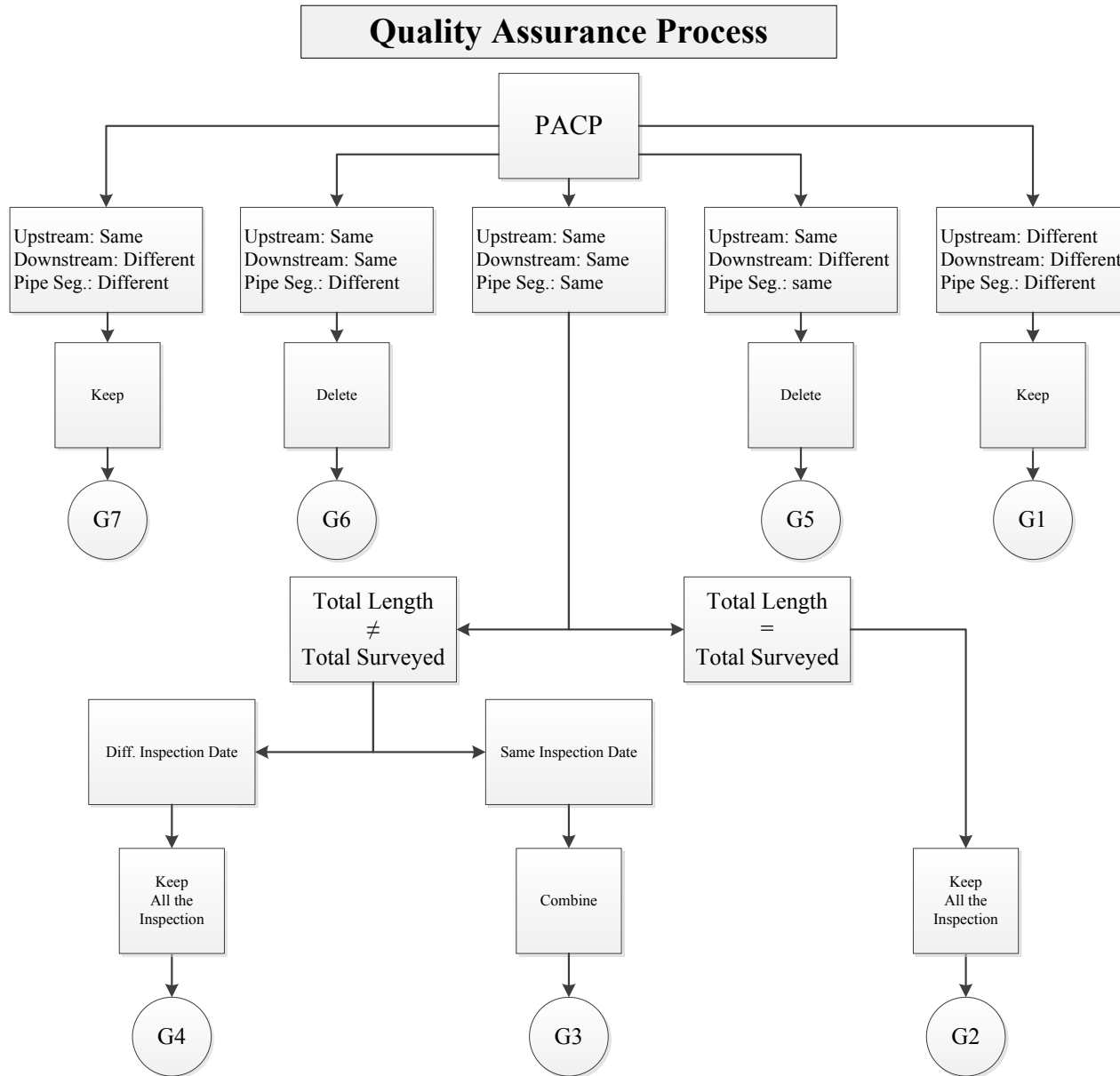
Application of Data Quality Assurance for Duplicates

- Second Database (SUD)
- Inspections:212
- Duplicate Inspections 52
- Deleted Inspection:2 (Instead of 52)

QA Groups	Number
G2	3
G3	23 (46/2)
G4	1
Deleted	2

PACP Grading System

- The PACP rating system focuses on the *structural* and *operational* condition of sewer pipes.

- Defects are classified into four different families
  - ➤ Structural
  - ➤ Operational
  - ➤ Construction
  - ➤ Other

- Defects are graded from 1 to 5 based on the severity of each defect.

PACP Grading System

- The Pipe Ratings Index (PRI) is the average of the grades within a pipe.

$$PRI = \frac{\sum Grades\ in\ the\ Pipe}{\sum Defects\ in\ the\ Pipes}$$

➢Defects with various criteria

➢Defects with grades "zero"

Defects with Grades "Zero"

Inspection 2658 FW

PRI 5 to PRI 2.5

ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_D	Value_2nd_D	Value_Percen	Joint
48763	2658	0	0	AMH					
48764	2658	0	0	MWL				25	
48765	2658	7	12	DAGS	S01			5	
48766	2658	7.7	40	DAE				5	✔
48767	2658	14.5	74	TFA		6			
48768	2658	35.1	157	RPR					
48769	2658	50.4	252	RPR					
48770	2658	65.6	532	H					
48771	2658	70.5	592	DAE				10	✔
48772	2658	81	670	HSV					✔
48773	2658	85.9	796	DAE				15	✔
48774	2658	92.5	834	TFA		6			
48775	2658	94.6	860	TFA		6			
48776	2658	100.2	916	DAE				10	✔
48777	2658	122.8	965	MWL				5	
48778	2658	130.3	1022	DAE				10	✔
48779	2658	172.3	1115	TFA		6			
48780	2658	174.7	1143	TFA		6			
48781	2658	175	1190	DSZ				15	
48782	2658	185	1270	DAE				5	✔
48783	2658	187.3	1296	MWL				40	
48784	2658	191	1307	DAGS	F01			5	
48785	2658	212.1	1342	TFA		6			
48786	2658	216.8	1375	AMH					

PRI Comparison
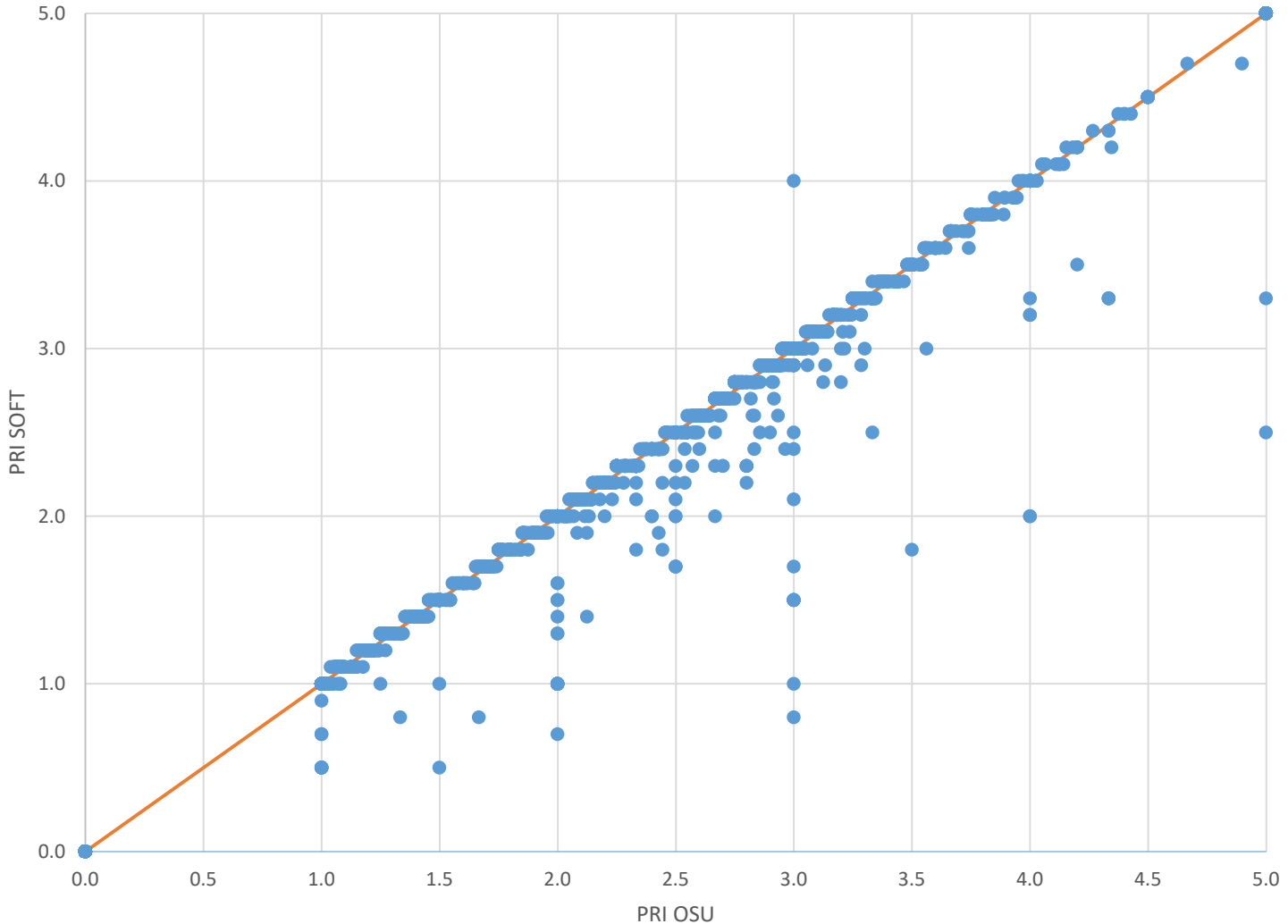
Conclusion

- In order to have an accurate evaluation of the sewer infrastructure condition across the nation, it is required to develop a high quality national sewer inventory.

- Data collection protocols were developed to convince sewer data owners to participate in One-Voice by granting access to their data.

- Moreover, a new data quality assurance process was developed to address the issues within the data and prepare the dataset for integration into the final inventory.

- The PACP grading system, which is widely used to determine pipe conditions in a sewer network, was evaluated in order to avoid inconsistencies in the calculation of ratings.

- To develop a One-Voice prototype database, more sewer data will be collected. After implementation of the new QA process, the new data will be integrated into the prototype database.

Thank You!

Q & A