Data Quality Evaluation Processes for PACP Inspection Data

Hossein Khaleghian
hossein.khaleghian@okstate.edu

Yongwei Shan, Ph.D., P.E.
yongwei.shan@okstate.edu

ONE VOICE SEWER SYSTEM

2017
INFRASTRUCTURE
REPORT CARD
ASCE

Wastewater

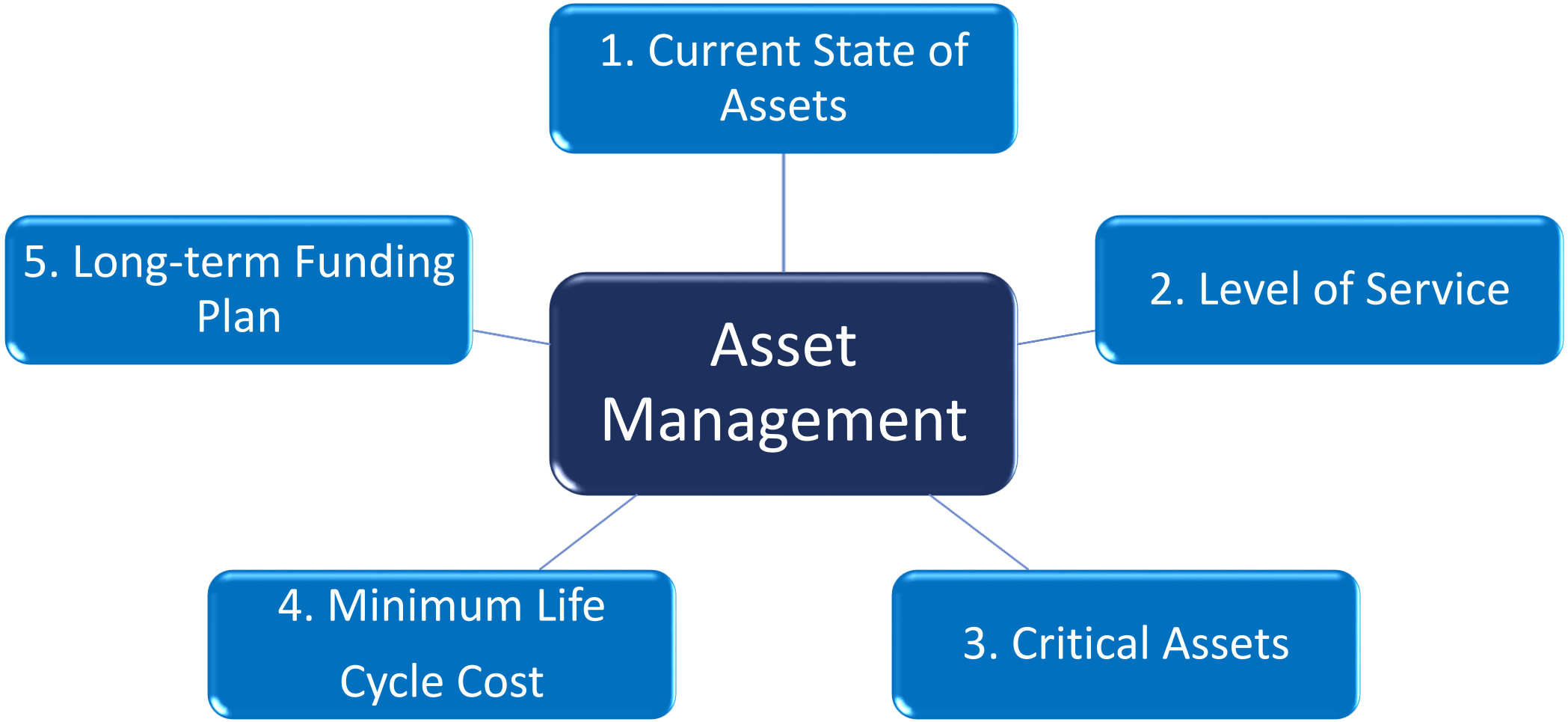
D+

Demand on treatment plants will grow more than 23% by 2032

Why Asset Management Plan (AMP)

- Increasing your knowledge of your system
- Prioritizing rehabilitation and replacement needs
- Reducing system “down-time” and the number of emergency repairs
- Showing investors and the public that you are using their money effectively and efficiently

Five Core Questions



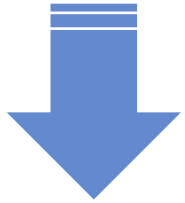
Condition Assessment

The outputs of inspections are raw data that need to be translated into information.

By analyzing the data, the condition of the asset is obtained.

In-depth review and analysis of data

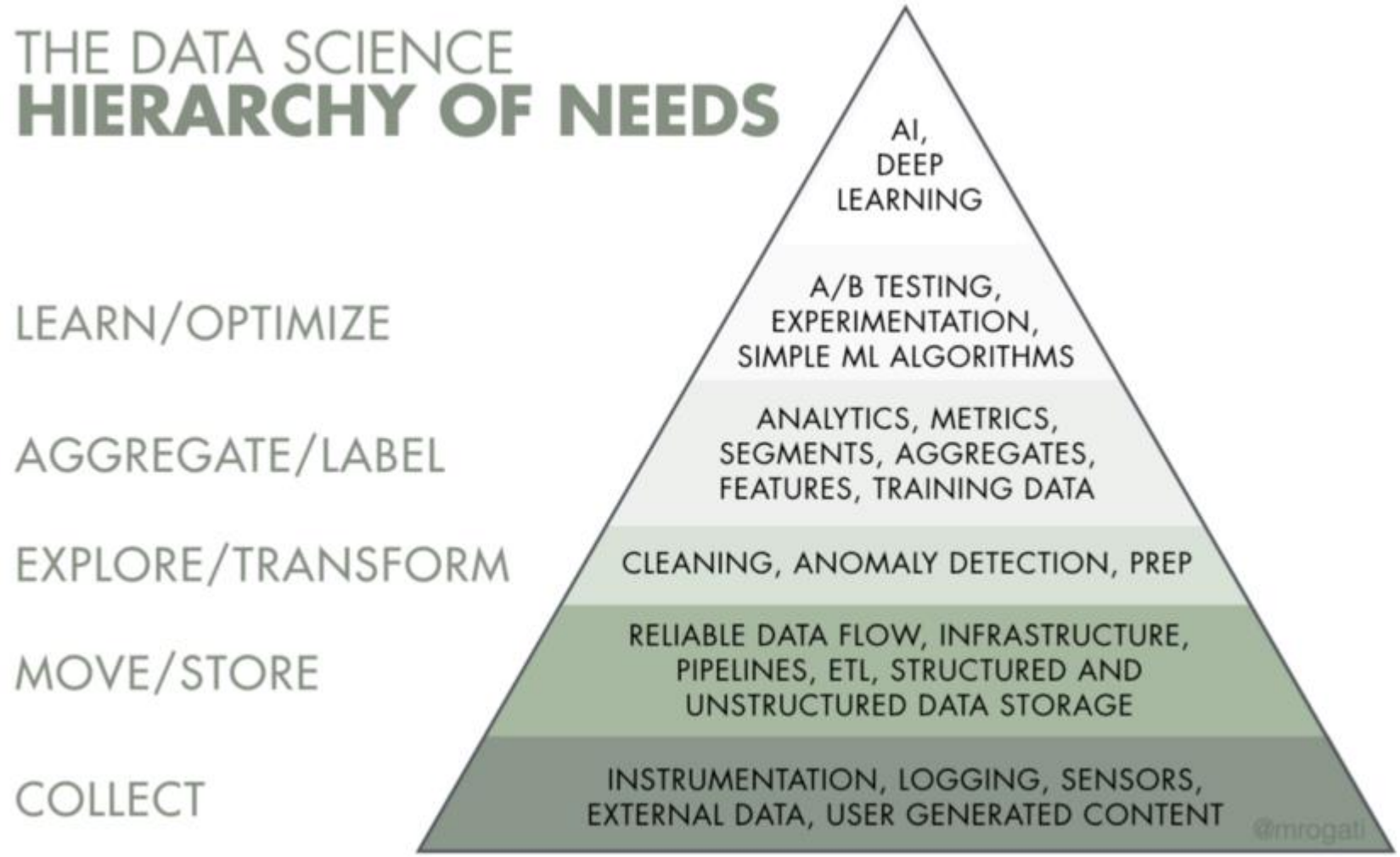
POOR DATA



POOR DECISIONS

INCREASED COSTS

THE DATA SCIENCE HIERARCHY OF NEEDS



Data Format (PACP)

InspectionID	Surveyed_By	Certificate_Num	Own	Custom	Drainage_Ar	PO_Numb	Pipe_Segment_Refere		Date	Time	Stre	Location_Deta
1	PDR	U-1104-1701				515267	K07 119	X	20120306	8:23	1232 Ki	
2	PDR	U-1104-1701				515267	K07 117	X	20120306	8:41	1232 Ki	
3	PDR	U-1104-1701				513080	L46 002	X	20120307	11:37	Honey	
4	PDR	U-1104-1701				513080	L46 003	H	20120307	11:53	Honey	
5	PDR	U-1104-1701				513080	L46 001	H	20120307	12:18	Honey	
6	PDR	U-1104-1701				513080	L46 004	H	20120308	11:34	Honey	
7	PDR	U-1104-1701				513080	L46 005	H	20120308	13:17	Cook R	
8	PDR	U-1104-1701					K19 116	X	20120309	8:35	1804 M	

ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_D	Value_2nd_I	Value_Perce	Joint	Clock_At_Fr	C
1	1	0	0	AMH					☐		
2	1	0	0	MWL				10	☐		
3	1	2.9	52	RMJ	S01			15	☑	7	
4	1	2.9	69	CM	S02				☑	7	
5	1	16.5	208	RBB				55	☐	2	
6	1	37	291	TBD		6			☐	2	
7	1	37	306	RBL				55	☐	2	
8	1	37	313	RMC				25	☐	2	
9	1	37.3	347	TBA		6			☐	9	
10	1	37.3	365	RFC					☐	9	
11	1	45	399	TFC		6			☐	9	
12	1	52	425	RMJ	F01			15	☑	7	
13	1	52	435	TBA		6			☐	1	
14	1	52	443	RFC					☐	1	
15	1	52.5	458	FM					☑	12	
16	1	63	487	RFJ					☑	12	

Data Format (Misc.)

ML_ID	Creation_	Last_Moc	US_MH	DS_MH	ML_Nam	Section_Length	Street	Location	Location_Details
1		12:19:14 PM	A001	T012	133951	355.53	LAKE O' ODONELL RD	Easement/Right of Wa	
2		10:16:40 AM	A002	A001	133950	92.42	Easement Ball Park Rd.		
3		12:37:51 PM	A003	A002	128457	188.04	UNIVERSITY ESMT	Light Highway	
4		1:02:24 PM	A004	A003	128456	219.87	SOLACE M FREEMAN	Main Highway - Urban	

MLI_ID	ML_ID	Inspection_Date	Inspection_Name	Creation_Date	Template_N	Last_Modifi	Start_Time	End_Tir
2	1	3/23/2011	76	3/23/2011 3:55:00 PM		14 12:19:14 PM	3:55:00 PM	011 9:14:
3	2	1/12/2005	74			14 10:16:40 AM	3:15:39 PM	
5	3	3/22/2011	71	3/22/2011 2:19:00 PM		014 2:00:54 PM	2:19:00 PM	
6	4	3/23/2011	84	3/23/2011 11:34:00 AM		14 12:00:00 PM	11:34:00 AM	
7	5	3/23/2011	85	3/23/2011 11:03:00 AM		14 11:59:44 AM	11:03:00 AM	
8	6	3/23/2011	86	3/23/2011 9:38:00 AM		14 11:59:21 AM	9:38:00 AM	

MLO_ID	MLI_ID	Distance	Digital_Time	VCR_Time	Code	Observation	Value_1st_D	Value_2nd_I	Value_Perce	Joi
15	2	0	43		AMH	Access Point M			0	
16	2	6.8	10.06		Mwl	Water Level			55	
17	2	155.2	435.24		TS	Tap Saddle	4		0	
18	2	175.3	490.24		TS	Tap Saddle	4		0	
19	2	364.6	902		AMH	Access Point M			0	
20	2	358.6	863.29		MWL	Water Level			5	
21	3	8.6			MWL	Water Level			15	
22	3	56.1			MWL	Water Level			25	
23	3	89.8			MWL	Water Level			30	

Incompatibility Problems

Data cannot imported from another format

Data cannot exported into different formats

6500 Inspections

85,000 conditions records

47,000 incompatibility errors

Data Transformation Tool

Provided Data

S_ID	S_Project_ID	S_Mandanto	S_Print_ID	S_Sat_ID	S_AutoNumb
FDA33F963A6C}	8BA162405386}	1210667AB892}	54457B3C5B53}		4
-115DDEFD173}	8BA162405386}	1210667AB892}	6E3F8E8C503E}		3
CE9A4350487C}	8BA162405386}	1210667AB892}	573B75594C3D}		1
CA0592F8135B}	8BA162405386}	1210667AB892}	506A99444863}		2
14FE28ADABD8}	8BA162405386}	1210667AB892}	4753645B4D88}		5

SI_ID	SI_Section_ID	SI_Mandantc	SI_Sat_ID	SI_Drawing_I	SI_Text_ID
158DECB983A3}	CE9A4350487C}	1210667AB892}			
9F541AB75A22}	-115DDEFD173}	1210667AB892}			
12103F76254A}	FDA33F963A6C}	1210667AB892}			
C4A94EE660DF}	14FE28ADABD8}	1210667AB892}			
307293C5CE7D}	CA0592F8135B}	1210667AB892}			

SO_ID	SO_Inspecs_I	SO_Sat_ID	SO_AutoNum	SO_Counter	SO_Position
DA5C45EBCAE}	C4A94EE660DF}		66	3	2.8042
340FBFF7BBA1}	158DECB983A3}		20	18	56.2966
C2F6ED6A32F0}	158DECB983A3}		7	5	19.5072
398EE71EB061}	158DECB983A3}		10	8	32.6136
CDF62626CDC}	158DECB983A3}		11	9	43.3121

Converted Data to PACP Format

InspectionID	Surveyed_By	Certificate_Number	Owner	Customer	Drainage_Area	PO_Number	Pipe_Segment_Reference	Date	Time	Street	City	Location_Details	Upstream_MH
566	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	566	2014-06-25	07:25:43			NULL	1043
567	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	567	2014-10-13	15:09:16			NULL	1045
568	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	568	2014-10-14	09:41:17			NULL	1047
578	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	578	2014-10-31	04:49:52			NULL	1067
593	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	593	2014-11-07	05:49:09			NULL	1099
597	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	597	2014-11-20	04:31:40			NULL	1108
598	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	598	2014-11-20	05:24:46			NULL	1110
599	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	599	2014-11-21	05:34:00			NULL	1112
600	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	600	2014-11-21	07:06:03			NULL	1114
601	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	601	2014-11-21	07:23:14			NULL	1116
602	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	602	2014-11-21	09:17:07			NULL	1118
609	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	609	2014-12-04	04:40:17			NULL	1126
610	Raul Alfaro	NULL	NULL	NULL	NULL	NULL	610	2014-12-04	05:52:55			NULL	1128

ConditionID	InspectionID	Distance	Counter	PACP_Code	Continuous	Value_1st_Dimension	Value_2nd_Dimension	Value_Percent	Joint	Clock_At_From	Clock_To	Remarks
5038	695	0	NULL	START AGAINST FLOW	NULL	NULL	NULL	NULL	0	NULL	NULL	NULL
5039	695	401.4	NULL	STOP	NULL	NULL	NULL	NULL	0	NULL	NULL	NULL
5040	695	66.5	NULL	Lateral	NULL	NULL	NULL	NULL	0	9	9	Left Tap
5041	695	72.6	NULL	Lateral	NULL	NULL	NULL	NULL	0	12	12	Right tap
5042	695	94.1	NULL	Lateral	NULL	NULL	NULL	NULL	0	10	10	Left Tap
5043	695	147	NULL	Lateral	NULL	NULL	NULL	NULL	0	2	2	Righ tap
5044	695	165.9	NULL	Lateral	NULL	NULL	NULL	NULL	0	10	10	Left Tap
5045	695	262.1	NULL	Pipe Type	NULL	NULL	NULL	NULL	0	12	6	Pipe change from sewer to water pipe.
5046	695	264.7	NULL	Lateral	NULL	NULL	NULL	NULL	0	3	3	Right tap
5047	695	284.2	NULL	Lateral	NULL	NULL	NULL	NULL	0	9	9	Left tap
5048	695	323	NULL	Pipe Type	NULL	NULL	NULL	NULL	0	12	6	Pipe changes back from water to sewer.
5049	695	343.7	NULL	Manhole	NULL	NULL	NULL	NULL	0	12	12	NULL
5050	695	343.8	NULL	General Observation	NULL	NULL	NULL	NULL	0	12	6	raeched M/H which starts with blue pipe.
5051	695	366.8	NULL	Water Level	NULL	NULL	NULL	NULL	0	6	6	Water almost half pipe.

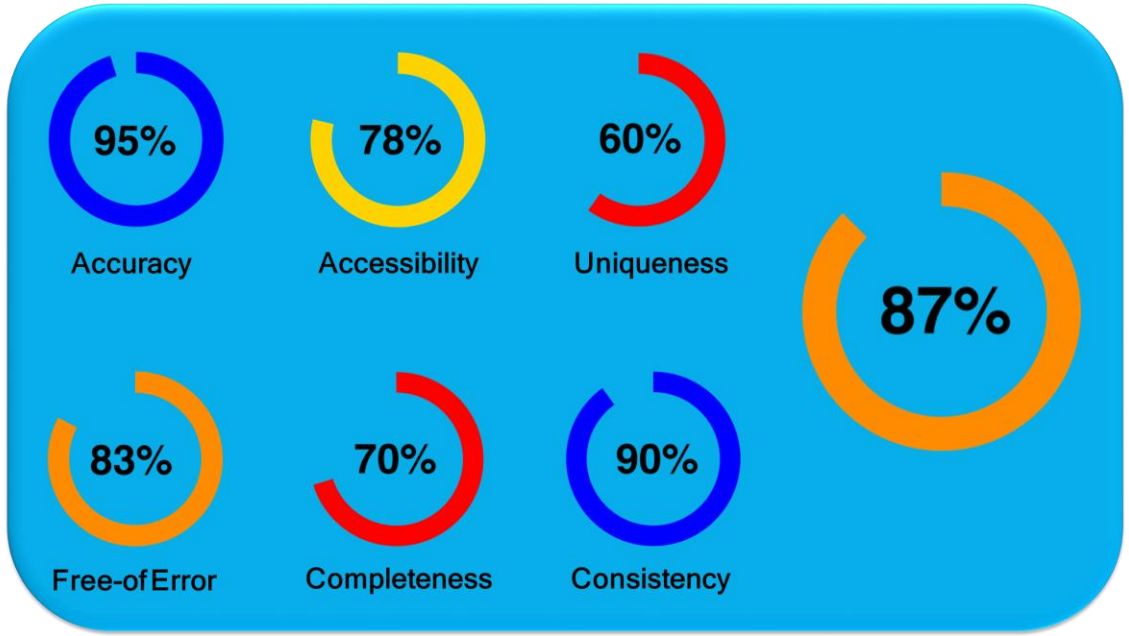
Benefits of Transformation Tool

- Intuitive user interface
- Ability to integrate with other applications already in use
- Configurability
- Scoring Assets
- Analytics to prioritize asset management activities
- Customizable Reporting

Data Quality Problems

- Problems with data that deserve review:

  - Missing Information
  - Redundancy
  - Accuracy
  - Inconsistency



ASSET								
KEY	SEGMENTID	UPSTREAM_MANHOLE	DOWNSTREAM_MANHOLE	PIPE_TYPE	PIPE_SHAPE	HEIGHT	WIDTH	ADDRESS
566	Sta. Maria & Water M/H.	1043	1044	PVC	Circular	8	8	Sta. Maria & Water

MAIN_INSPECTION							
KEY	ASSET	REVERSED	SCHEDULED_DATE	WEATHER	COMMENT	PROJECT	OPERATOR
695	566	-1	6/25/2014 7:15:33 AM	Dry	Started of on Sta. Maria towards Juarez upstream towards Water.	86	Raul Alfaro

MANHOLE	
KEY	MANHOLE_ID
1043	Juarez M/H..
1044	Water & Sta. Maria
1045	Sta Maria M/H

OBSERVATION							
DISTANCE	CODE	LENGTH	CLOCK_FROM	CLOCK_TO	SEVERITY	INSPECTION	COMMENT
0	START AGAINST FLOW					695	

ASSET Table

- ASSET Table provides the pipe features (72 rows):
  - Manhole Information
  - Material
  - Shape
  - Dimension
  - Length

Column	Description	Comment	Completeness	Consistency	Uniqueness	Accuracy
KEY	Primary Key		100%	N/A	100%	N/A
SEGMENTID	Developed based on MAHONHOLE_ID	Data Redundancy	100%	40%	40%	26%
UPSTREAM MANHOLE	Manhole number Ref to MANHOLE Table	This number is not representing unique MH	100%	?	?	?
DOWNSTREAM MANHOLE	Manhole number Ref to MANHOLE Table	This number is not representing unique MH	100%	?	?	?
WIDTH	The width is not required for circular pipes (all pipes in the database)	Data Redundancy	100%	N/A	0%	N/A
ADDRESS	Based on SEGMENTID	Data Redundancy	100%	30%	0%	94%
ASSET_LENGTH	The total length of the pipe	For all the pipes the asset length is equal to the surveyed footage	83%	N/A	N/A	0%

MAIN_INSPECTION Table

- MAIN_INSPECTION Table provides some information on the inspection (72 rows):
  - Operator
  - Weather
  - Reason
  - Date
  - Comments

Column	Description	Comment	Completeness	Consistency	Uniqueness	Accuracy
KEY	Primary Key	The inspection Identifier	100%	N/A	100%	N/A
ASSET	Foreign Key to ASSET Table	It helps to join two tables	100%	100%	100%	100%
REVERSED	It shows the reverse setup	Because of manhole numbering issues it could not be evaluated	100%	N/A	N/A	N/A
COMMENT	Direction and Location	Data Redundancy. The useful information is hard to be extracted.	100%	70%	18%	87%
OPERATOR	The name of the inspector	There is only two names	100%	N/A	N/A	N/A
REASON	The purpose of the inspection	9 different reasons has be detected	98%	N/A	N/A	N/A
SURVEYED FOOTAGE	The length of the surveyed segment	The information comes from observations	98%	100%	N/A	100%
DATE_START DATE_END	The date and time of inspection	DATE_START will be considered as a reference	100%	100%	N/A	100%

MANHOLE Table

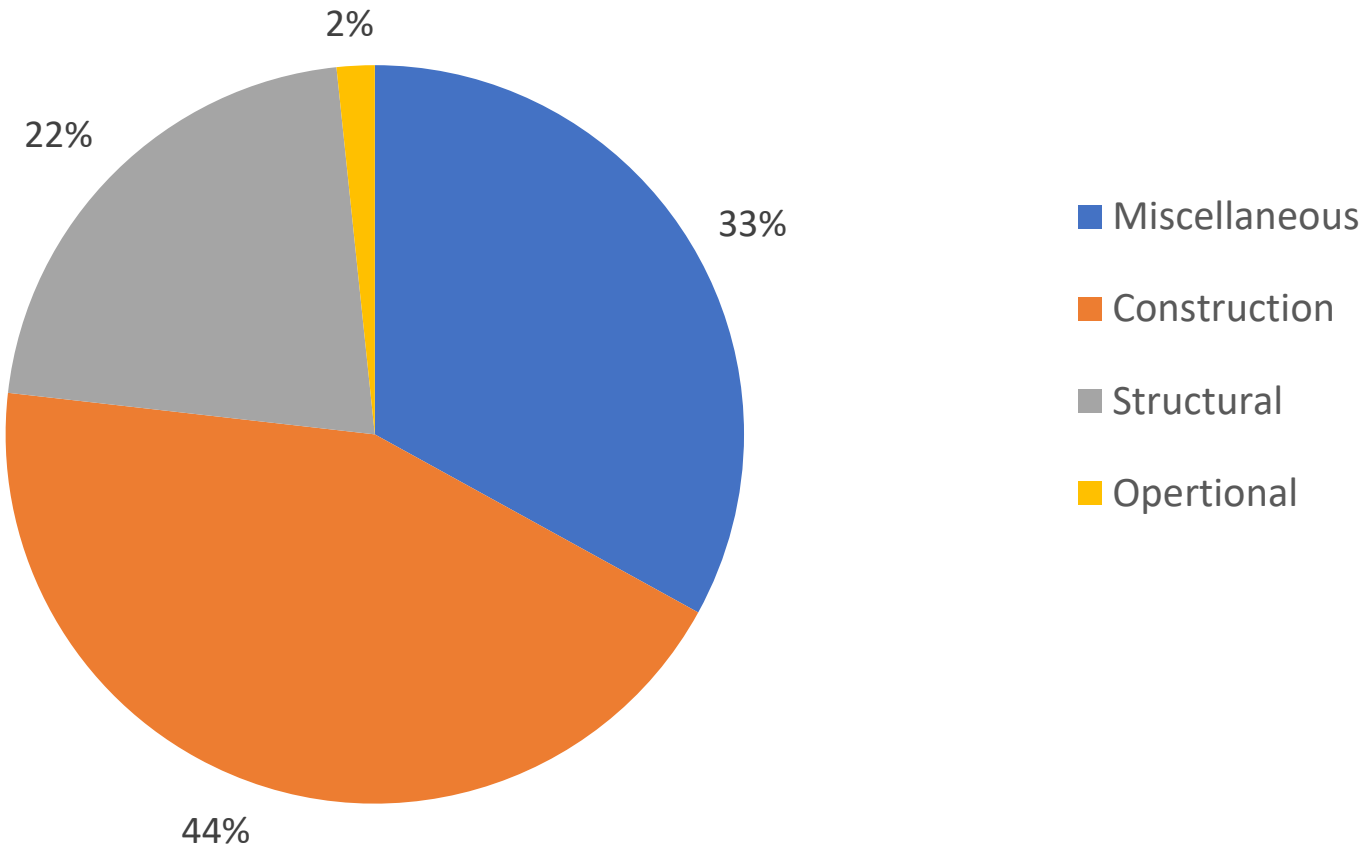
- MANHOLE Table provides some information on the manholes (143 rows):
  - Manhole ID
  - Address

Column	Description	Comment	Completeness	Consistency	Uniqueness	Accuracy
KEY	Primary Key	The keys are unique but they are not manhole identifier	100%	N/A	100%	N/A
MANHOLE_ID	It represents the manhole address	The manholes are not defined individually	100%	75%	40%	N/A

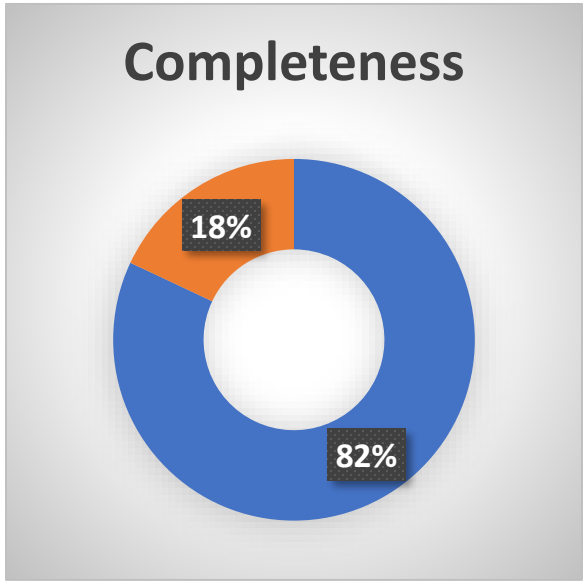
OBSERVATION Table

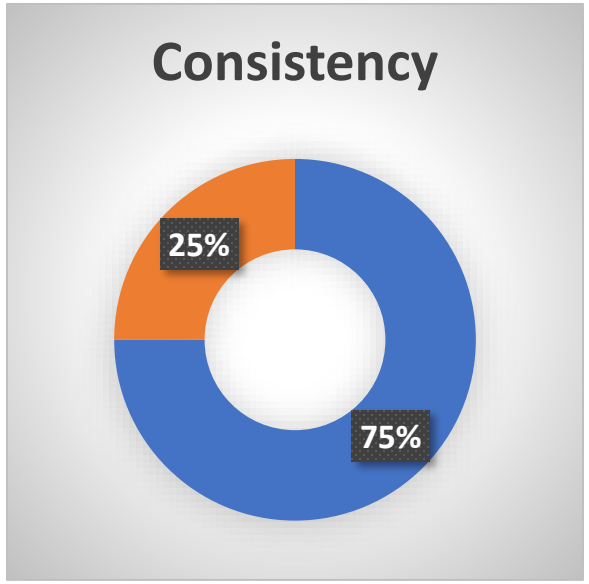
- OBSERVATION Table provides some information on the conditions (724 rows):
  - Code
  - Distance
  - Clock Positions
  - Severity

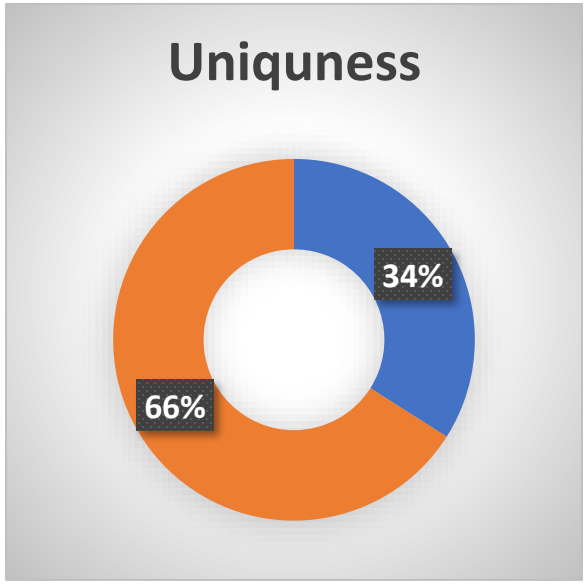
Defect Groups



Data Quality of the Defects







What to Do?

- Data should be evaluated at the first step!
- Develop algorithms to address the data quality problems.
  - Data Redundancy
  - Extracting information
  - Cross checking the values
- This process can be done automatically or by the operator.

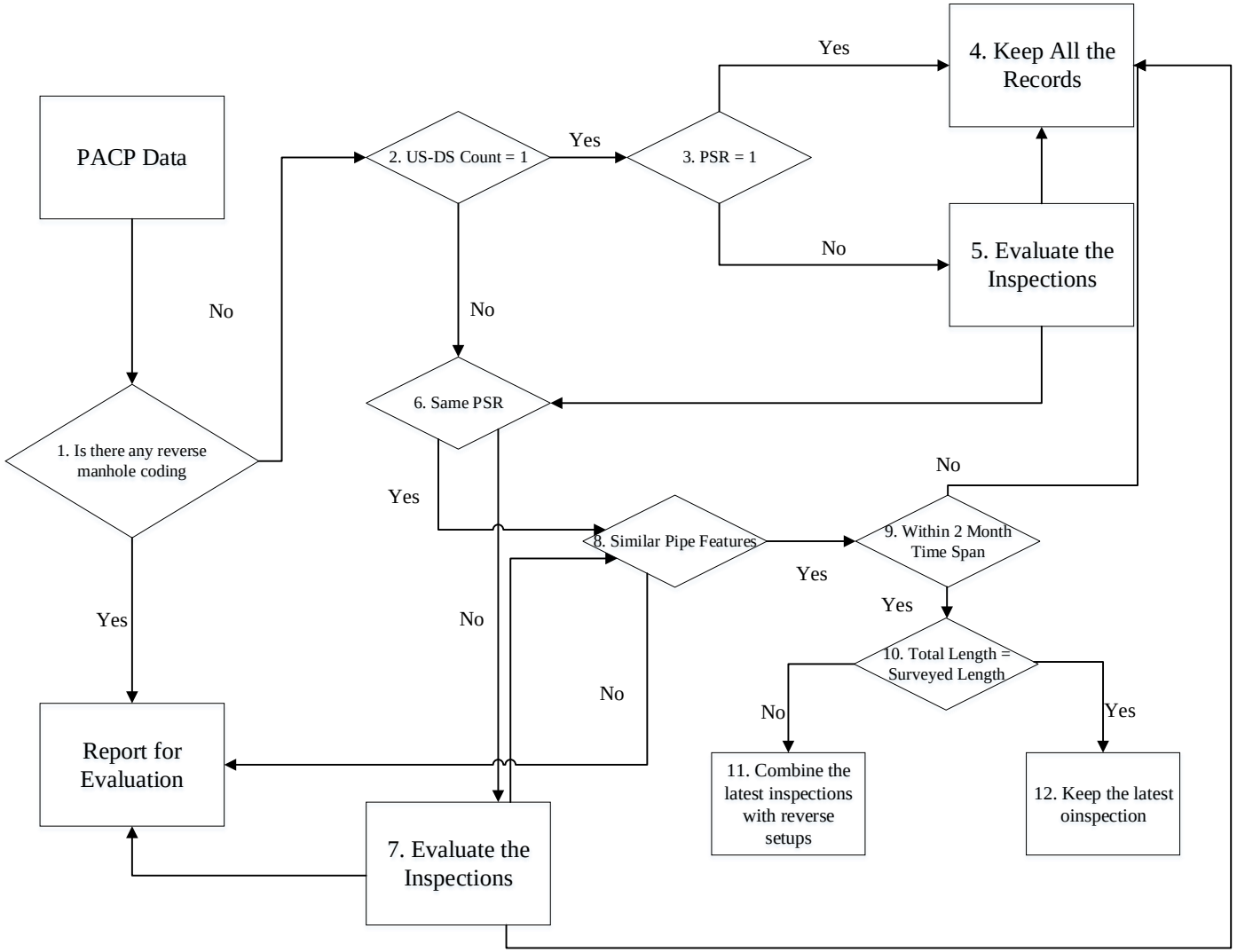
INSPECTION

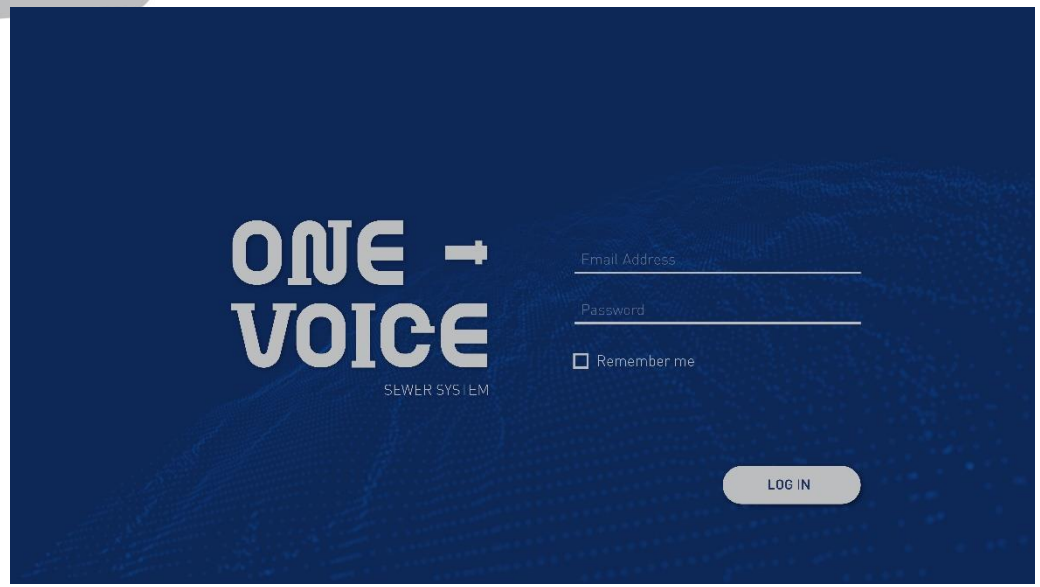
Inspection	Date & Time	Asset	UPMH	DSMH	Address	Shape	Length	Operator	Reason	Weather
2306	5/8/2018 4:43 AM	1919	3494	3495	Iturbide & Convent M/H	Circular	35.9	Raul Alfaro	Regular	Dry
Comment	Started from Iturbide & Convent downstream towards Lincoln									

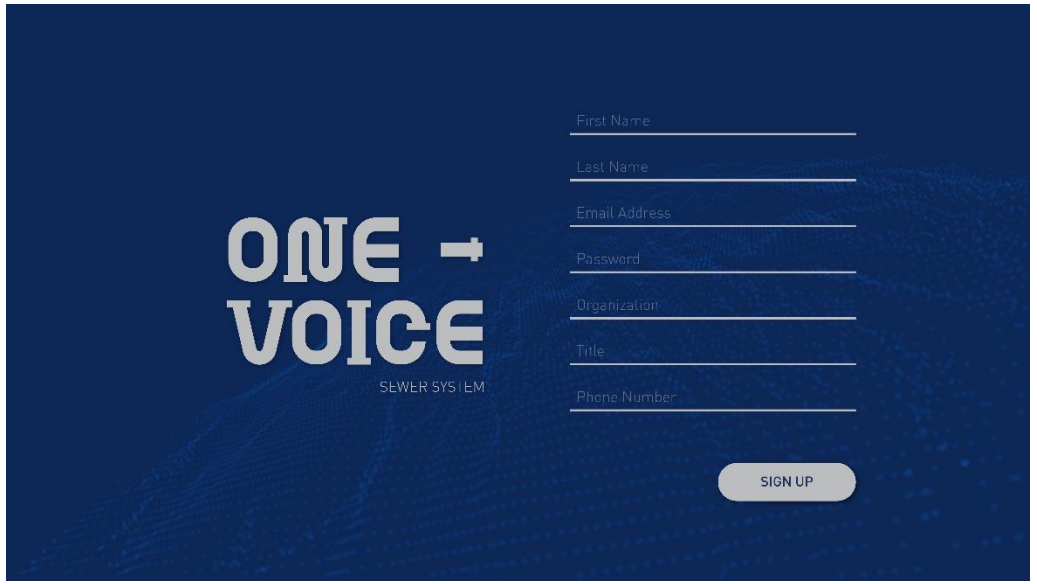
OBSERVATIONS

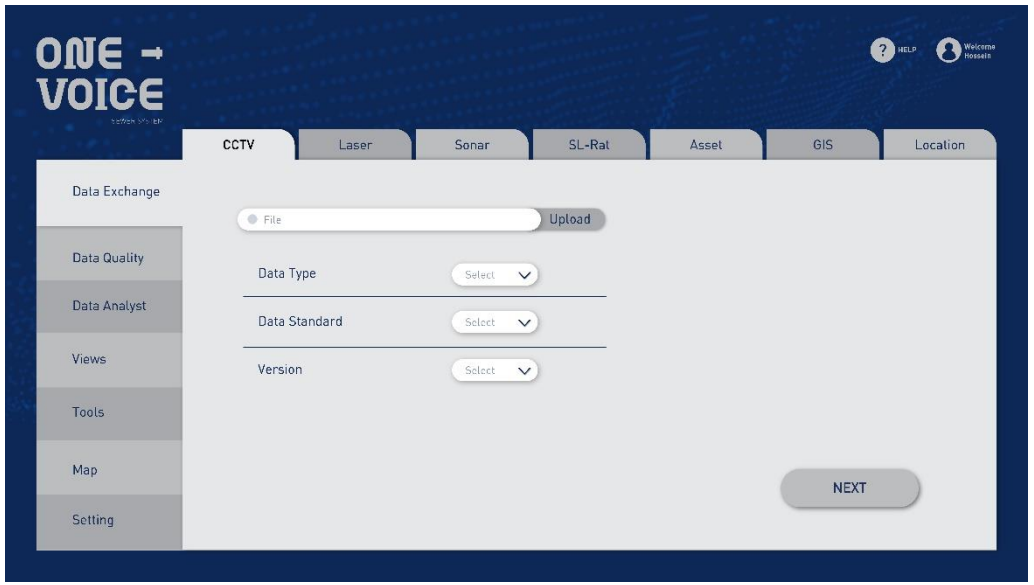
Key	Distance	Code	Clock_From	Clock_To	Severity	Comment
15664	0	Start with flow				
15665	4	Pipe Type				
15666	11.8	Broken	12	12	Hole	
15667	13.6	Pipe Type				
15668	35.9	General Observation				Line turns left unable to continue
15669	35.9	Abandoned Survey				
15670	35.9	Stop				

Automated Data Quality Resolution




ONE VOICE
SEWER SYSTEM
Email Address
Password
Remember me
LOG IN


ONE VOICE
SEWER SYSTEM
First Name
Last Name
Email Address
Password
Organization
Title
Phone Number
SIGN UP


ONE VOICE
HELP
Welcome Hossein
CCTV
Laser
Sonar
SL-Rat
Asset
GIS
Location
Data Exchange
Data Quality
Data Analyst
Views
Tools
Map
Setting
File
Upload
Data Type
Select
Data Standard
Select
Version
Select
NEXT


ONE VOICE
HELP
Welcome Hossein
Dashboard
Metrics
Duplicate
Error Correction
Evaluation
Data Exchange
Data Quality
Data Analyst
Views
Tools
Map
Setting
87.0%
Accuracy
30.0%
Accessibility
40.0%
Uniqueness
85.0%
Total Score
50.0%
Free of Error
80.0%
Completeness
90.0%
Consistency

Conclusion

- In-Depth Data Review is essential
- Data Structure can be improve for data accessibility
- Quality problems can be addressed automatically
- Data analytics can be implemented on quality approved databases

ONE VOICE SEWER SYSTEM