responsive.

HIMSS 15

Annual Conference & Exhibition

APRIL 12–16, 2015
Chicago, IL

Realizing Value. **Transforming Health.**

The Power of EHR Data Forensics: Knowing What You've Got
April 14, 2015

Lehigh Valley Health Network
A PASSION FOR BETTER MEDICINE.

Lori Yackanicz, Administrator Clinical Informatics, Lehigh Valley Health Network

Joy Ales, Senior Consultant, Encore, A Quintiles Company

encore
A Quintiles Company

www.himssconference.org

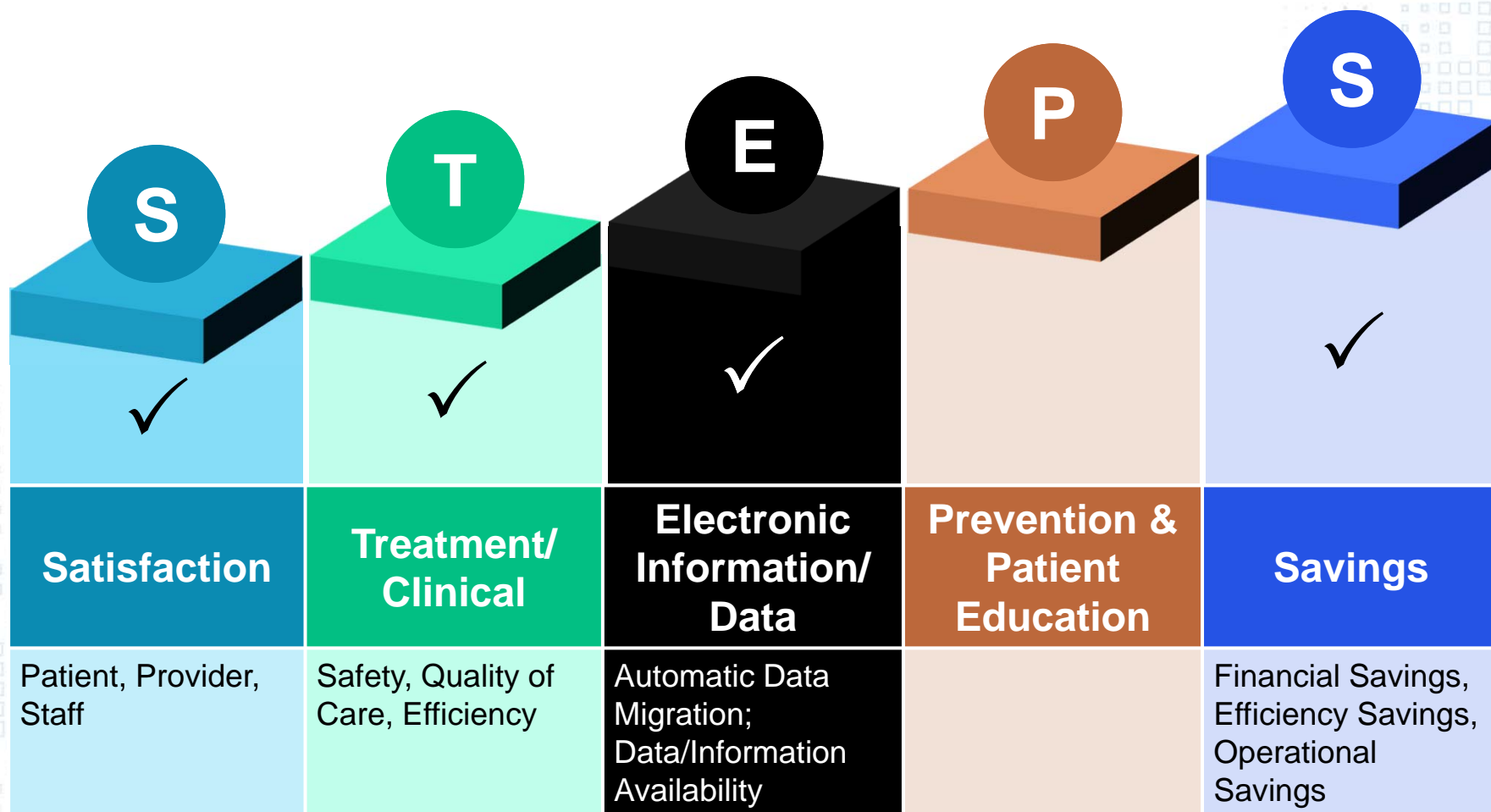# Conflict of Interest

Lori Yackanicz, MBA

Joy Ales, MHA, BSN, RN

- Have no real or apparent conflicts of interest to report.

Learning Objectives

- Define data forensics

- List benefits of data forensics

- Describe the types of projects that benefit from data forensics

- Explain what types of resources are needed for a successful data forensics project

An Introduction to the Benefits Realized for the Value of Health IT

Satisfaction	Treatment/ Clinical	Electronic Information/ Data	Prevention & Patient Education	Savings
Patient, Provider, Staff	Safety, Quality of Care, Efficiency	Automatic Data Migration; Data/Information Availability		Financial Savings, Efficiency Savings, Operational Savings

Lehigh Valley Health Network

Who We Are

| Cedar Crest | 17th Street | Muhlenberg | Health Centers |

- Largest academic community hospital in PA
- Largest Level 1 Trauma Center in region
- Certified Stroke Center
- Employees – 9,800
- Medical Staff – 1,200+
- Nurses – 2,334

- Magnet Hospital
- 163,000 ED visits
- 68,602 admissions
- 981 acute care beds
- 3 hospital campuses
- Revenues over $1 Billion

HIMSS 15

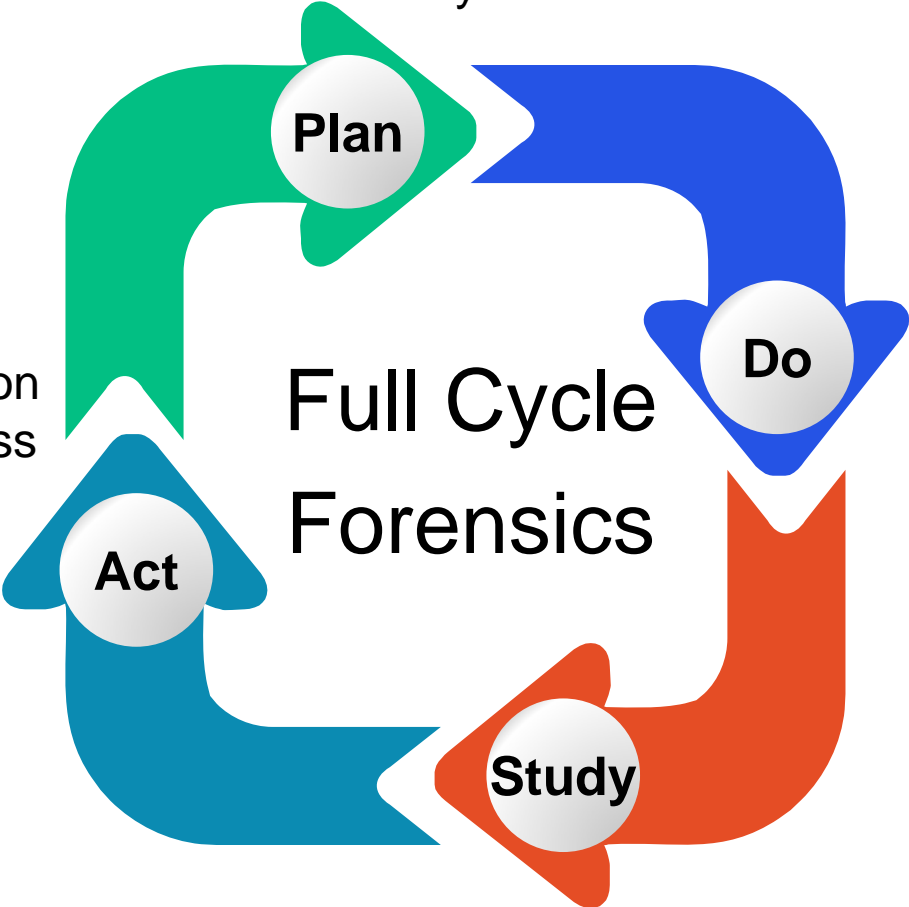# Data Forensics Is the Diagnostics that Supports Effective Re-use of Your Data

Understanding the Current State of Your Data

MAYO CLINIC

Data Forensics Cycle

Prioritize data elements & source systems.

Plan

Do

Full Cycle Forensics

Conduct quantitative & qualitative analysis of data. Review & validate findings with SMEs.

Decide remediation efforts &/or process changes.

Act

Study

Analyze captured profile information and categorize data (High, Medium, Low reliability).
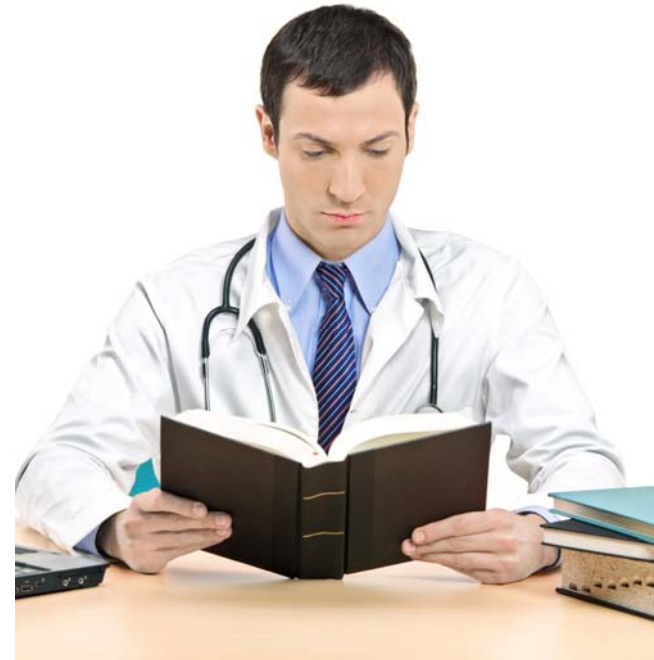
HIMSS 15

Quantitative Assessment: Gather the Facts

Diagnostics and Physical Assessment of the Data

Assess the Facts

Ranges

Validity

Data Integrity

Complete-ness

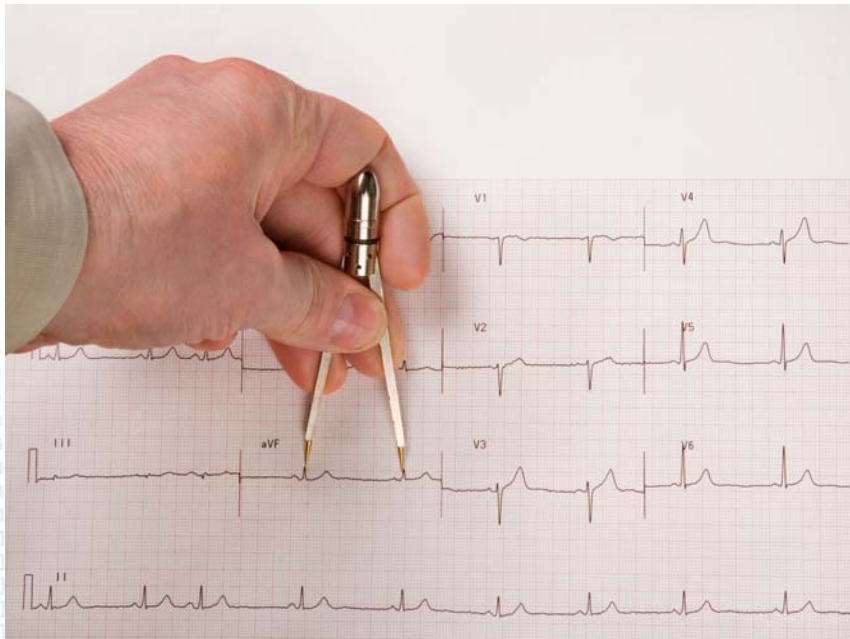Frequency

Patterns

Understand the Expected

Data Specifications
Business Rules
Data Standards
Metadata

Qualitative Assessment: Evaluate the Results & Diagnose the Problems

Measure Data Accuracy

Evaluate Consistency & Synchronization

Data Findings Classification

Facts: Expected & Actual — **Logic, Assessment, & Meaning**

Quantitative	Qualitative
Duplicate data values	Data field definition (intent)
Null values	Field use compliance with intent
Field length (min & max)	Drop down list individualization
Format compliance with expected	Hijacked field identification
System default values expected	Data reliability categorization
Mandatory Field Indicator	Comments & assessment of data
Data Patterns	Clinically duplicative field identification
Reference table compliance with expected	Recommendations/considerations

Data Findings Classification

Illustration

	Facts: Expected & Actual	Logic, Assessment, & Meaning
Field	**Quantitative**	**Qualitative**
Temperature (F)	Numeric; nn.nn – nnn.nn	Values below and above what is humanly possible (e.g. negative values, over 500)
Medication Dose	Text field; reference table compliance; null values %	Values outside of appropriate clinical range (e.g. 0.5 mg, 5 ml)

Qualitative Data: State of your Data

High

Probability of consistent, reliable data

Medium

Probability of consistent, reliable data
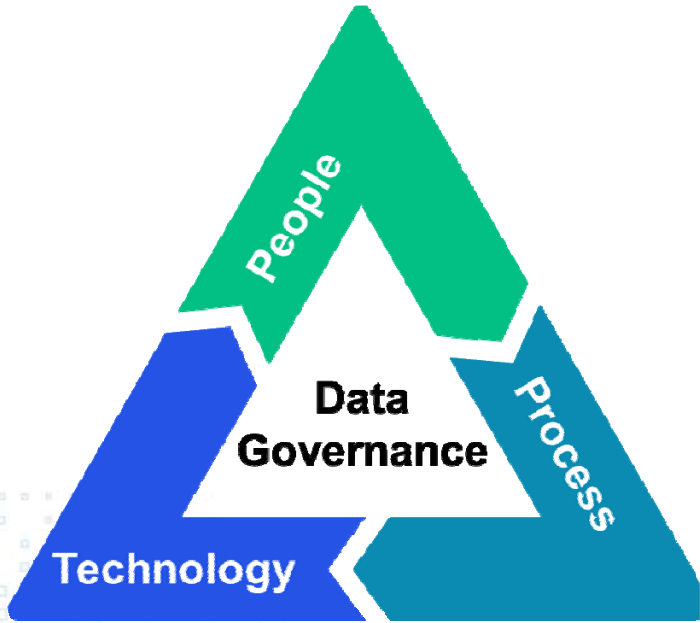- May require mapping of multiple code sets to a common code set

Low

Probability of consistent, reliable data
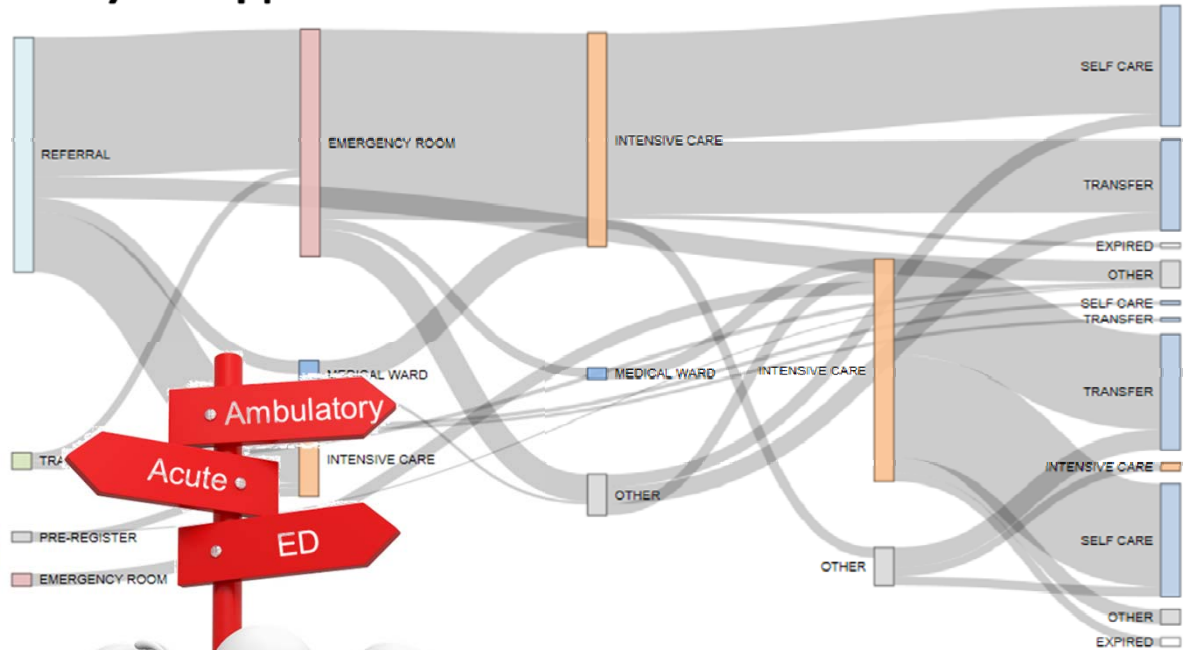- Multiple types of data captured in the same field

Where Do You Apply Data Forensics?

Data Governance — People, Process, Technology

Analytics Applications

Data Interoperability

What's the Problem?

Planning to migrate 25 legacy applications to a single EHR in 2015

| Our EHR vendor did not want to migrate poor quality data | We had concerns about potential hijacked fields | We didn't know our data quality | We didn't know which source data we could/ should migrate |

Our Objective…

…was to examine the data from the legacy EHR systems being converted to the new EHR, and to identify any technical and/or business issues with the data.

Focused on three major EHR systems:
- Ambulatory System
- Acute Care System
- Emergency Department System

Example 1

Ambulatory System Sample Data Sets Reviewed

Data Subject Area	Number of Distinct Data Elements
Immunizations (Influenza & Pneumonia)	26
Smoking & Tobacco Use	20
Vital Signs	12
Height/Weight/Head Circumference	8

Example 1:

Data Findings (Smoking)

EMR TABLE ELEMENT NAME	PAS CIG SMOK	SMOK HX TOTA	SMOK HX PPD	SMOK YR ST	SMOKSTARTAGE	QUIT SMK STG
Expected						
EMR FIELD DESCRIPTION	Passive smoke exposure	Pack Years	Packs/day smoking	Year started smoking	Age started smoking	Smoke cessation stage
EXPECTED DATA TYPE FORMAT	Text string AA or AAA	Numeric x.xx-xx	Numeric Text string	Numeric XXXX	Numeric x or xx	Text string
FIELD EXPECTED DUPLICATE (Y/N)	Y	Y	Y	Y	Y	Y
ALLOWABLE CHARACTER LENGTH	2000	2000	2000	2000	2000	2000
FIELD REFERENCE TABLE (Y/N)	Y	N	Y	N	Y	Y
FIELD DEFAULT VALUES (Y/N)	N	N	N	N	N	N
MANDATORY (Y/N)	N	N	N	N	N	N
Data Facts						
SAMPLE RECORD COUNT	21665	7409	10956	3308	35	4393
OBSERVED DATA TYPE FORMAT	Numeric Text string	Numeric Text string	Numeric Text string	Numeric Text string	Numeric Text string	Numeric Text string
DUPLICATE (Y/N)	Y	Y	Y	Y	Y	Y
MIN LENGTH	1	1	1	1	2	2
MAX LENGTH	65	73	69	77	18	40
NULL COUNT	12	31	58	1	0	0
% NULL	0.1%	0.4%	0.5%	0.0%	0.0%	0.0%
REFERENCE TABLE VALUES NON-COMPLIANCE COUNT	267	n/a	8071	n/a	9	51
REFERENCE TABLE VALUES NON-COMPLIANCE %	1.2%	n/a	73.7%	n/a	25.7%	1.2%
NOT WITHIN FORMAT COUNT	245	5116	n/a	1239	9	n/a
% NOT WITHIN FORMAT	1.1%	69.1%	n/a	37.5%	25.7%	n/a
HIJACKED FIELD (Y/N)	N	N	N	N	N	N
COMMENTS			Reference table different for EMR form A and B.	Data captured is alphanumeric. Would expect to see year started (e.g. 1965).		The field description is different for each form and the reference table values differ for each form.

Example 1

Expected Data Findings (Smoking)

EMR TABLE ELEMENT NAME	SMOK HX TOTA
EMR FIELD DESCRIPTION	Pack Years
EXPECTED DATA TYPE FORMAT	Numeric x.xx – xx
FIELD EXPECTED DUPLICATE (Y/N)	Y
ALLOWABLE CHARACTER LENGTH	2000
FIELD REFERENCE TABLE (Y/N)	N
FIELD DEFAULT VALUES (Y/N)	N
MANDATORY (Y/N)	N

Expected

Example 1

Data Facts (Smoking)

EMR TABLE ELEMENT NAME	SMOK HX TOTA
SAMPLE RECORD COUNT	7409
OBSERVED DATA TYPE FORMAT	Numeric Text string
DUPLICATE (Y/N)	Y
MIN LENGTH	1
MAX LENGTH	73
NULL COUNT	31
% NULL	0.4%
REFERENCE TABLE VALUES NON-COMPLIANCE COUNT	N/A
REFERENCE TABLE VALUES NON-COMPLIANCE %	N/A
NOT WITHIN FORMAT COUNT	5116
% NOT WITHIN FORMAT	69.1%
HIJACKED FIELD (Y/N)	N
COMMENTS	

Data Facts

Full Cycle Forensics
Plan Do Act Study

Example 1

Summary Findings: High Reliability
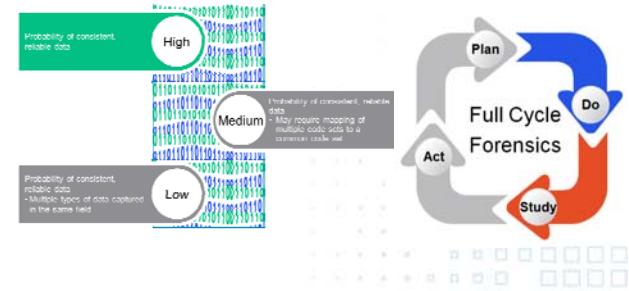
Data Element Name	Quantitative	Qualitative
HEIGHT	Numeric* & text string; Range: 2.64-73174.6	Less than 0.01% of values are clinically improbable.
HEIGHT (CM)	Numeric*& text string; Range: 6.7-185863.48	Less than 0.01% of values are clinically improbable.
WEIGHT (KG)	Numeric* & text string; Range: 0.14-12176.82	Less than 0.01% of values are clinically improbable.
WEIGHT	Numeric* & text string; Range: 0.31-26789	Less than 0.01% of values are clinically improbable.

*This is a calculated field and has numeric restrictions for free text entry.

Example 1

Summary Findings: Medium Reliability
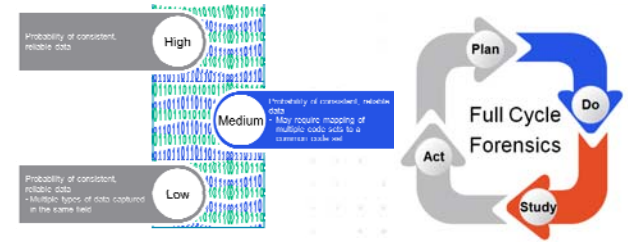
Data Field Description	Quantitative	Qualitative
SMOKSTATODAY SMOK STATUS TOBACCO USE SMOKES CIGARET SMKG	Numeric & text string; Compliance with reference table values range: 92-98%.	SMOKSTATODAY: interfaced field. Multiple fields capture similar clinical assessments. Can map observed values to common field values. SMOK STATUS data capture is most consistent.
ADV TO QUIT SMOK ADVICE	Text string; Compliance with reference table values range: 90%-99%.	Multiple fields capture similar clinical assessments. Can map observed values to common field values. ADV TO QUIT data capture is most consistent.

Example 1

Summary Findings: Low Reliability
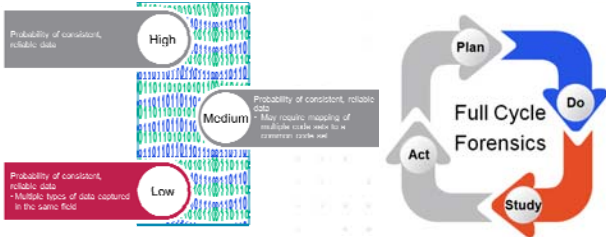
Data Element Name	Quantitative	Qualitative
CIGARPIPEUSE CIGARS WEEK TOBACCOTYPE ORALTOBACUSE	Numeric & text string; Free text fields except for TOBACCOTYPE which is free text and reference list field.	Multiple fields capture similar clinical assessments. Data capture is highly inconsistent.
SMOK HX PPD	Numeric & text string; 30-40% reference table compliance.	Data capture is highly inconsistent.
SMOK HX TOTA SMOKSTARTAGE SMOK YR ST	Numeric & text string; one field utilizes a reference table and the others utilize free text; 60-75% reference table compliance.	All fields capture similar clinical assessments. Data capture is highly inconsistent between fields.

Example 2
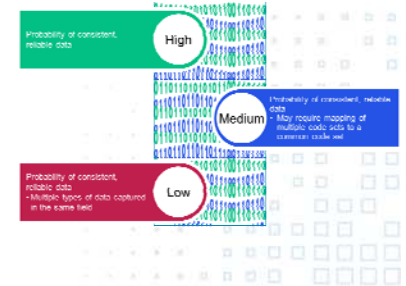
Multiple Legacy Systems

Acute Care   Ambulatory

Example 2: Multiple Legacy Systems

Allergies Data Set

EMR System	Number of Distinct Data Elements
Ambulatory EMR	15
Acute Care EMR	4
Total	**19**

Example 2: Multiple Legacy Systems Summary of Categories - Allergies

| Summary Categories - Allergies	Number of Distinct Data Elements
High Reliability	5
Medium Reliability	3
Low Reliability	5
Do Not Migrate	6
Total	**19**

Example 2: Multiple Legacy Systems

Allergies Summary Findings

There are five clinically significant allergy data fields
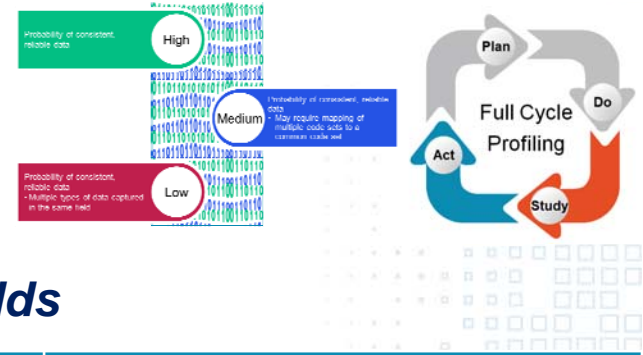
| Legacy App	Data Field Name	Field Definition	Data Reliability Category
Amb	NAME	The name of the allergen (e.g. Sulfa)	High 25%-30% non-coded
Amb	DESCRIPTION	The reaction to the allergen (e.g., rash, anaphylaxis)	Low Fee text field
Acute Care	DESC	The name of the allergen (e.g. Sulfa). Populated by a pick list.	High 0% non-coded
Acute Care	DESC_OVERRIDE _TXT	The name of the allergen (e.g. Sulfa). Populated by free text & historical allergens.	Low Free text field
Acute Care	MOD_TXT	The reaction to the allergen (e.g., rash, anaphylaxis)	Low Free text & drop down list

How Did We Benefit from Data Forensics?
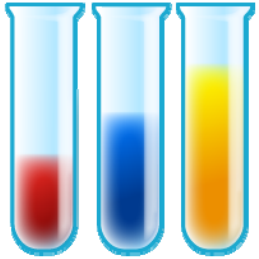
By having the data forensic analysis, we had the ammunition needed to support our decision to electronically migrate data to the new EHR.

The data forensic results provided us with the guidelines that we needed to support our data migration (such as coding special rules) to ensure that data presentation on the receiving EHR side was clean.
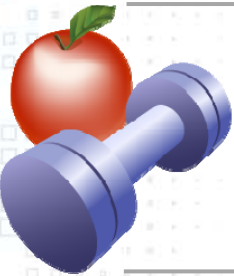
The prerequisite data forensics effort validated to the EHR vendor, as well as ourselves, which data was "healthy" enough to migrate.

Lessons Learned & Recommendations

Engage your SME's

Custom Screens

Individualized documentation screens can lead to inconsistent and duplicate data fields

Mandatory Fields

Some mandatory fields encourage poor data quality by creating workarounds

Business Need

Get started by identifying & defining the business cases needing quality data

Transactional Data

Not every data element is equally important

Automated Processes

Automated processes for data capture and documentation provide clean and consistent data

Standardization

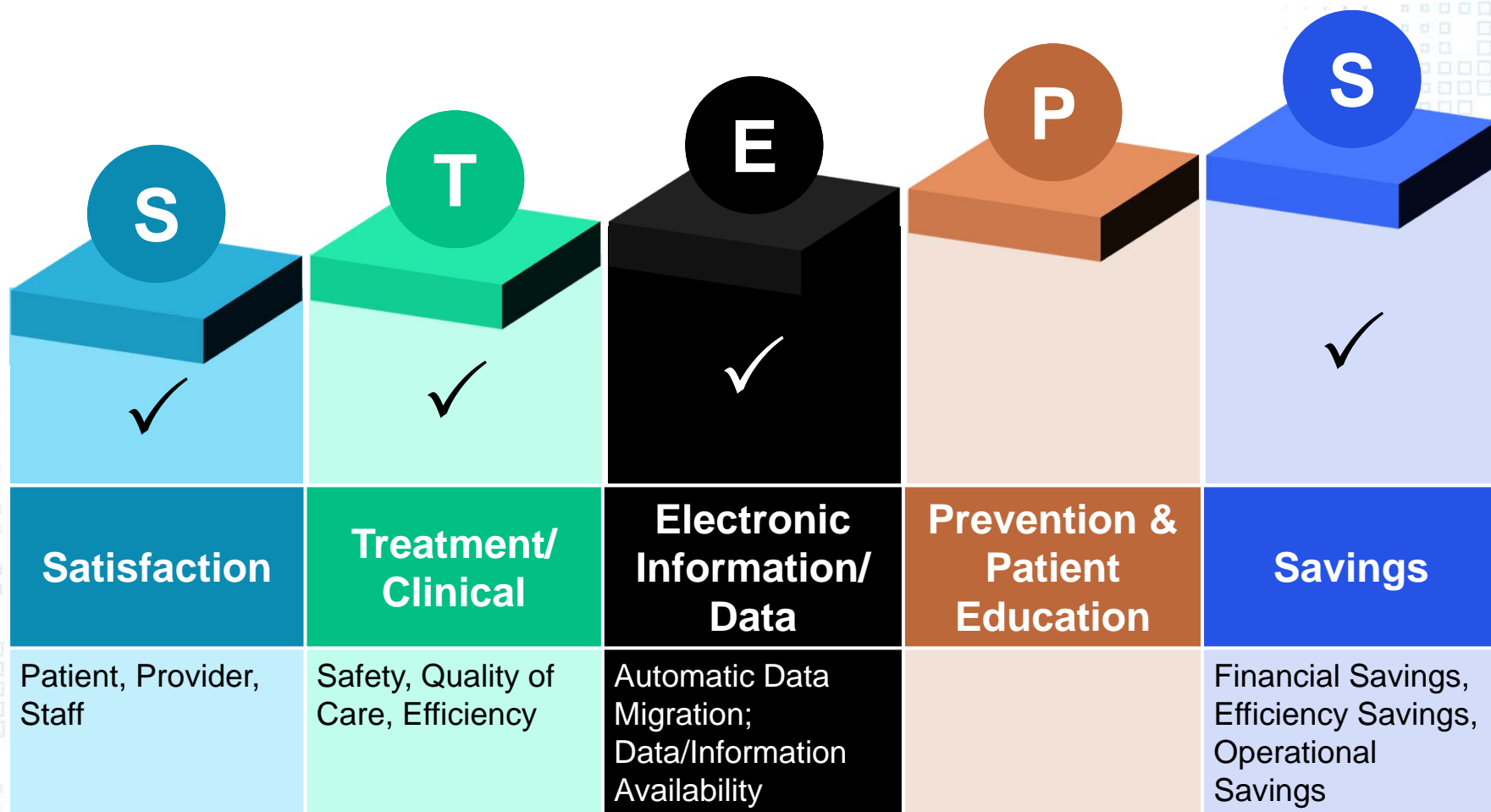Non-standardized drop-down lists can lead to inconsistent data

Workflow

Understand & build the workflow to prevent the barriers to good data capture (workarounds)

Monitoring

Build a process to continually monitor the quality of data

HiMSS13

An Introduction to the Benefits Realized for the Value of Health IT

S	T	E	P	S
Satisfaction	**Treatment/ Clinical**	**Electronic Information/ Data**	**Prevention & Patient Education**	**Savings**
Patient, Provider, Staff	Safety, Quality of Care, Efficiency	Automatic Data Migration; Data/Information Availability		Financial Savings, Efficiency Savings, Operational Savings

Questions

Lori Yackanicz
Lori.Yackanicz@lvhn.org

Joy Ales
Jales@encorehealthresources.com