



ALL IN DATA FOR COMMUNITY HEALTH

- Community Health Peer Learning Program
- Data Across Sectors for Health

Considerations for Development and Use of a Master Person Index (MPI)

July 26, 2016
3 - 4 pm EST

Presenters



Clare Tanner, PhD

Co-Director of Data Across Sectors for Health (DASH),



Melissa Moorehead

Policy Analyst and Project Manager, Michigan Public Health Institute



Stephen Singer, MCP

Senior Manager of Data Analytics, Camden Coalition of Healthcare Providers



Dan Chavez, MBA

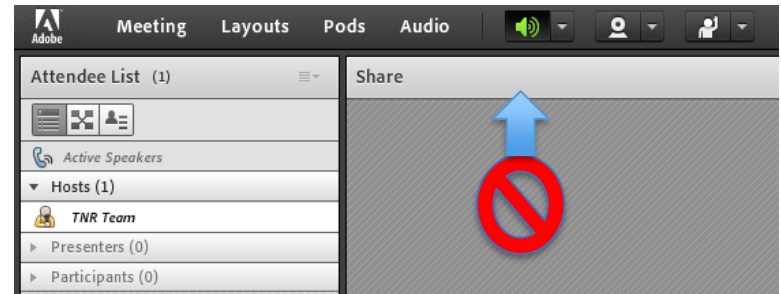
Executive Director, San Diego Health Connect

Meeting Information

- Meeting Link:
<http://academyhealth.adobeconnect.com/mpi/>
- Registered: Select “Enter with your login and password” and enter the following:
 - Username: [enter email address used to register for the webinar]
 - Password: index
 - Click “Enter Room”
- Unregistered Guest: Select “Enter as a guest” and enter your name, e.g., Kelsi Feltz, CHP.

Meeting Information

- Conference Line: 1-866-546-3377
- Access Code: 6478553818
- Reminders:
 - Please **hard-mute your computer speakers** and the **speakers in the web conference**
 - Please **mute your phone line** when you are not speaking to minimize background noise
- Technical difficulties? Email us at chpinfo@academyhealth.org



Chat Feature

- To share your comments using the chat feature:
 - Click in the chat box on the left side of your screen
 - Type into the dialog box and click the send button
- To signal to presenters you have a question / comment:
 - Click on the drop down menu near the person icon and choose *raise your hand*



Agenda

- Introduction (3 minutes)
 - Clare Tanner, DASH NPO, will provide a brief introduction to All In
- MPI Case Study #1 (12 minutes)
 - Stephen Singer, MCP, Senior Manager of Data Analytics at Camden Coalition of Healthcare Providers, will discuss how Camden Coalition uses and continues to evolve their person-level matching using various methodologies in the research settings.
- MPI Case Study #2 (12 minutes)
 - Daniel Chavez, MBA, Executive Director at San Diego Health Connect and a CHP Subject Matter Expert community, will discuss how San Diego Health Connect is using an HIE and addressing standards to improve automated patient matching capability.
- Discussion (30 minutes)
- Wrap-Up (3 minutes)

DASH and CHP are All In!

Community Health Peer Learning Program

- NPO: AcademyHealth, Washington D.C.
- Funded by the federal ONC
- 15 participant and subject matter expertise communities

Data Across Sectors for Health (DASH)

- NPO: Illinois Public Health Institute in partnership with the Michigan Public Health Institute
- Funded by the RWJF
- 10 grantee communities

All In: Data for Community Health



1. Support a movement acknowledging the social determinants of health



2. Build an evidence base for the field of multi-sector data integration to improve health



3. Utilize the power of peer learning and collaboration

Considerations & Questions about Record Linkage

&
MPI's



Camden
Coalition
of Healthcare Providers

Stephen Singer, Senior Program Manager,
Data Analytics & Quality Improvement

The Camden Coalition Data Environment

HIE

vendor-managed.
*MPI via
... a black box*

corrections

IDs &
events

user-customizable,
vendor-hosted.
*MPI via HIE linkage
+ deterministic linkage
+ extensive manual review*

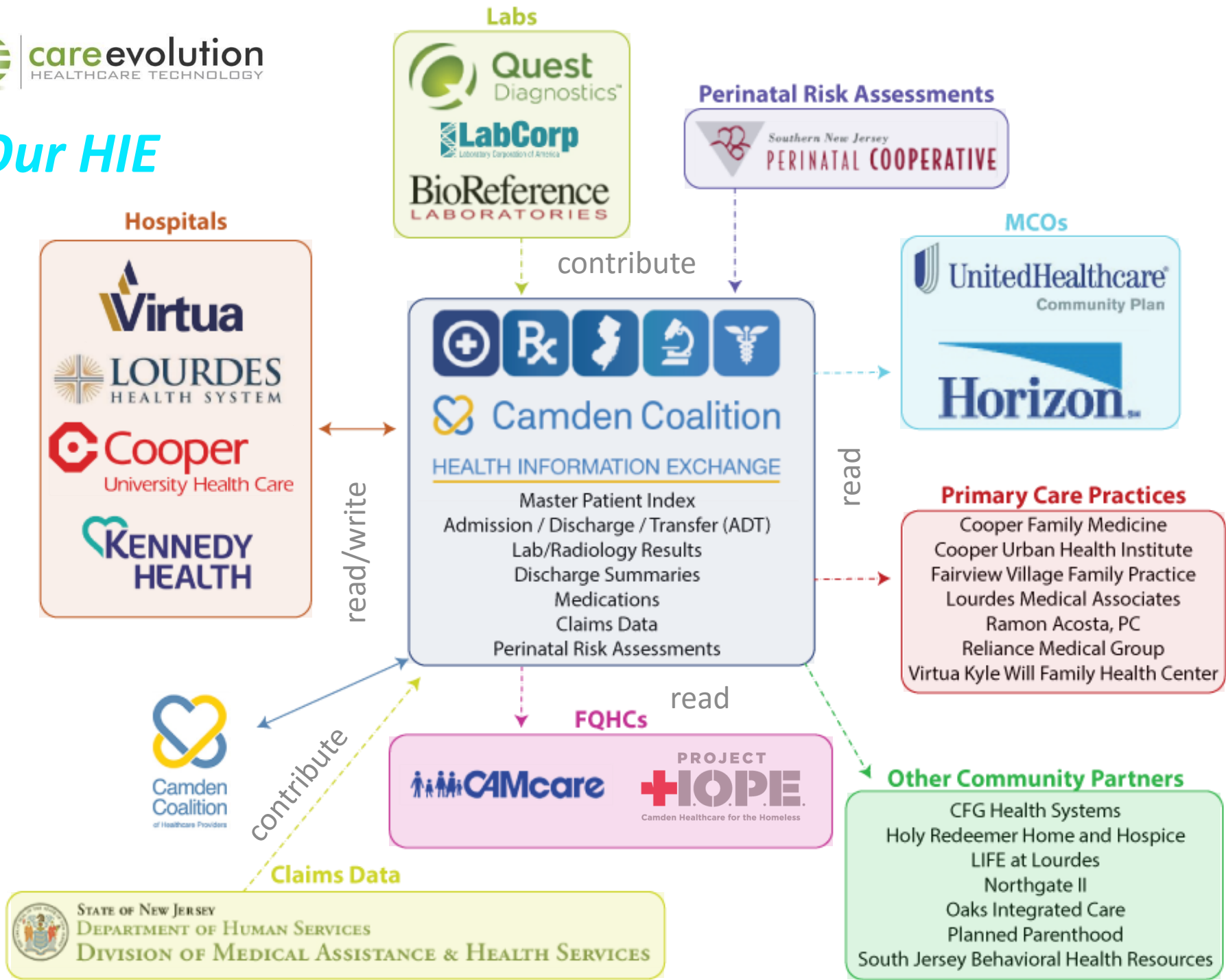
Internal performance & care
tracking

home-grown
PostgreSQL database.
*No MPI. previously
linked via commercial
probabilistic linkage
software, temporarily via
hierarchical, fuzzy,
deterministic match*

retrospective
me hospital claims

cross-sector
integrated data

Our HIE



Cross-sector Integrated Data “System”

Existing Data Sharing:

- 1) All-payor hospital claims from 4 regional health systems biannual (plus a 1 time extract from a 5th)
- 2) State Medicaid Claims monthly
- 3) Camden Police Department no fixed schedule (arrest, call-for-service, & overdose)
- 3) Camden City School District no fixed schedule (enrollment, truancy, absenteeism, & suspension data)
- 4) Camden County Jail (booking & release) monthly
- 5) NJ State Prison (booking & release) bi-monthly
- 6) property data (citywide vacancy survey) one time

In Discussion:

- 1) Homelessness Management Information System
- 2) State Mortality Records

Integrated Identifiers

	Hospital Claims	State Medicaid	School District	Police Arrest	Sate Prison	County Jail	HMIS	Death Cert.
First Name	Green	Green	Green	Green	Green	Green	Green	Green
Middle Name	Green	Green	Green	Green	Green	Green	Red	Green
Last Name	Green	Green	Green	Green	Green	Green	Green	Green
Name Suffix	Green	Green	Green	Green	Green	Green	Red	Green
Alias	Red	Red	Red	Red	Red	Red	Red	Green
Date of Birth	Green	Green	Green	Green	Green	Green	Green	Green
Date of Birth Alias	Red	Red	Red	Red	Red	Red	Red	Red
Date of Death	Yellow	Red	Red	Red	Red	Red	Red	Green
Gender	Green	Green	Green	Green	Green	Green	Green	Green
Race and/or Ethnicity	Green	Green	Green	Green	Green	Green	Red	Green
Street Address	Green	Green	Green	Green	Red	Red	Red	Green
Zip Code	Green	Green	Green	Red	Red	Red	Red	Green
City	Green	Green	Green	Green	Red	Red	Red	Green
State	Green	Green	Green	Green	Green	Green	Red	Green
County	Red	Green	Green	Red	Red	Green	Red	Green
SSN	Green	Red	Red	Red	Red	Red	Red	Green
MRN	Green	Red	Red	Red	Red	Red	Red	Red
Federal Bureau of Prisons #	Yellow	Red	Red	Red	Red	Red	Red	Red
State Bureau of ID #	Red	Red	Red	Green	Green	Green	Red	Red
State/Local Bureau of Criminal ID #	Red	Red	Red	Red	Red	Red	Red	Red
Inmate ID	Red	Red	Red	Green	Green	Green	Red	Red
Family ID	Red	Red	Green	Red	Red	Red	Yellow	Red
Family Members	Red	Red	Red	Red	Red	Red	Yellow	Green

Why

an

MPI?

To resolve existing data dis-integration (*linkage*) & prevent future data dis-integration (*data management*)

So that we can correctly identify & characterize patients for appropriate & coordinated care, accurate quality metrics, and research?

But

Some data are undecidably ambiguous.

(What about twins?)

New data require unstable IDs.

Data entry is only partially controllable.

Data entry isn't the only source of error.



Buy

vs

Build?

1. How soon? How fast?
2. How expensive?
(\$ + training + staff-hours)
3. How flexible & stable?
4. How interoperable?
5. How accountable?

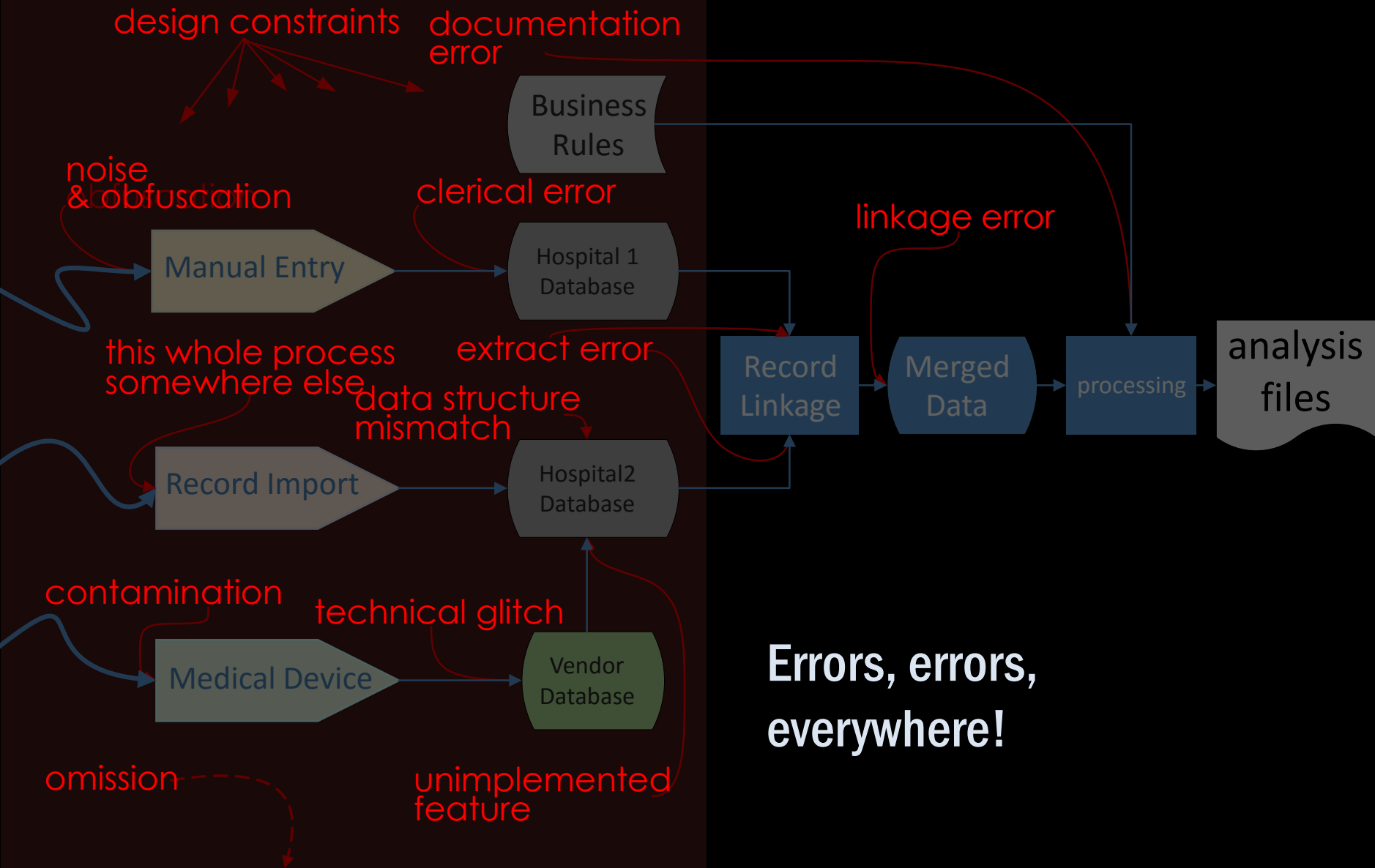
5 questions for any vendor:

1. Can I get ALL of my data back?
2. How do you do it?
3. Who can I talk to...
outside of sales and marketing?
4. How responsive is tech support?
5. Can you flag records by linkage quality?

		hospital	mrn	dob	last	first	mid	ssn						
A Real (extreme) Case	W		1	06/03/1965	SMITH	SIMON		296	1	4	6	4	1	1
			2	06/20/1965	HIGHSMITH	SIMON		296	1	4	6	5	1	1
			3		BEN		296	1	1	6	5	1	1	
			4		RUIZ	SIMON		296	1	4	6	4	1	1
X	5	BENN			N	296	1	4	6	5	1	0		
		6	06/20/1965	SMITH	LARRY		296	1	3	6	5	5	5	
Y		7	06/20/1965	RUIZ	BEN		L	296	1	3	6	8	4	4
		8		RUIZ		L	296	1	4	6	5	1	0	
		9		SYMON		N	296	1	7	5	5	6	1	
		10				N	296	1	4	6	5	1	0	
		11	06/20/1966	SMITH	LARRY		296	1	3	6	5	5	5	
		10					296	1	4	5	8	8	8	
Z		12	06/30/1966		JAMES			296	1	4	6	4	1	1
		13	06/20/1965	RUIZ				296	1	4	5	5	6	0
		14			BEN		296	1	7	6	4	1	1	
		13						296	1	7	6	5	1	1
		14			LARRY		J	296	1	4	6	5	1	1
		15				296	1	4	6	4	1	1		
		16		RUIZ-SMITH		N	296	1	4	6	4	1	1	
				SMITH			296	9	8	6	8	4	3	
	15	06/30/1965	RUIZ	SIMON		296	1	4	6	4	1	1		

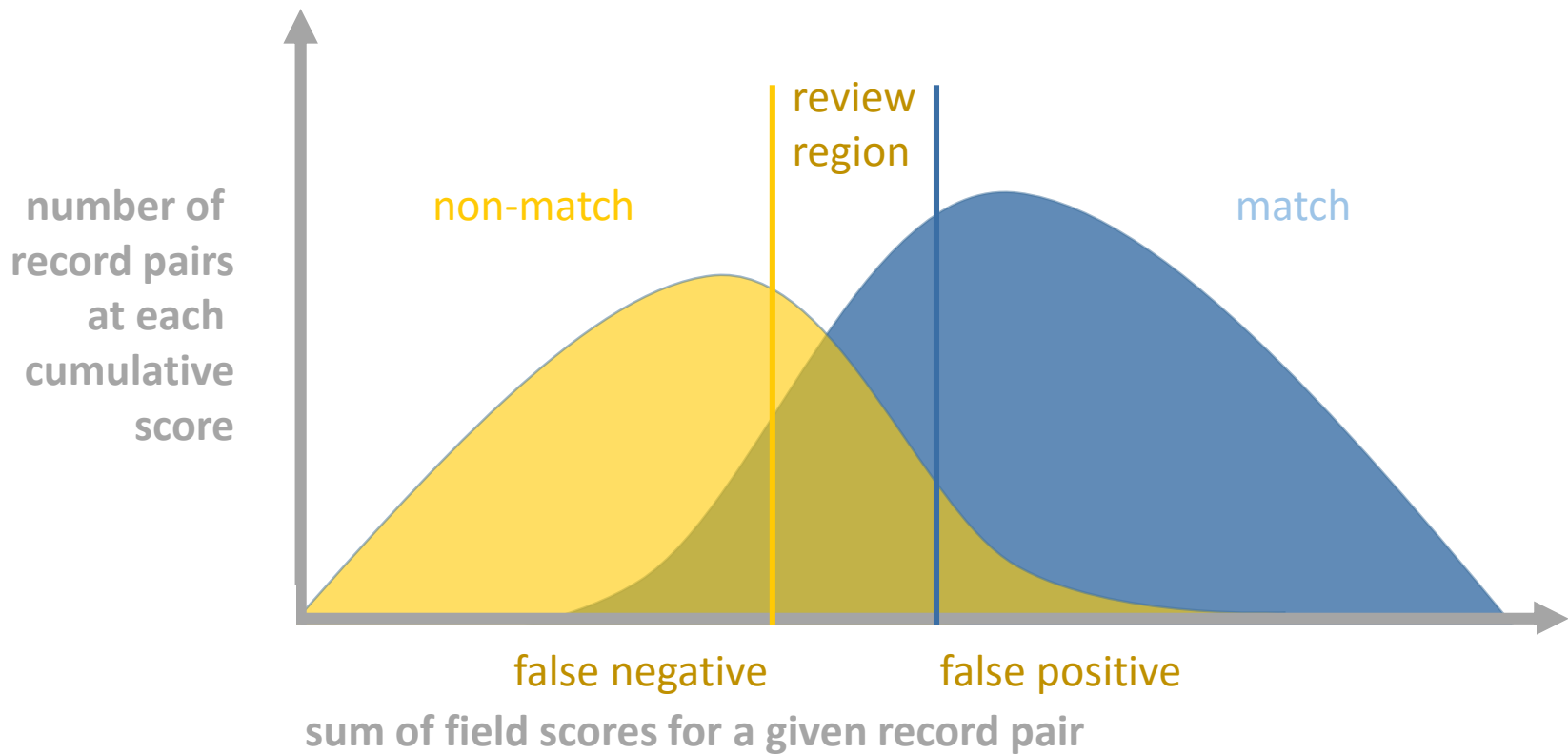
Data Production (out of your hands)

Data Manipulation



MRN	SSN	First name	Last name	Date of birth
2	296146511	WILLIAM	HIGHSMITH	6/20/1965
14	296146511	LARRY	RUIZ	6/20/1965
14	296176411	JOHN	RUIZ	6/20/1965
14	296176411	JOHN	RUIZ	6/20/1965
5	296146510	JON	RUIZ	6/20/1965
5	296146510	WILYAM	RUIZ	6/20/1965
16	296986843	LARRY	SMITH	6/20/1965
16	296146411	LARRY	RUIZ-SMITH	6/20/1965
16	296146411	LARRY	SMITH	6/20/1965
12	296146411	JAMES	RUIZ	6/20/1966
15	296146411	WILLIAM	RUIZ	6/30/1965
15	296146411	WILLIAM	RUIZ	6/20/1966
15	296146415	WILLIAM	RUIZ	6/20/1967

Deterministic linkage groups together records that are equal on subsets of identifier fields

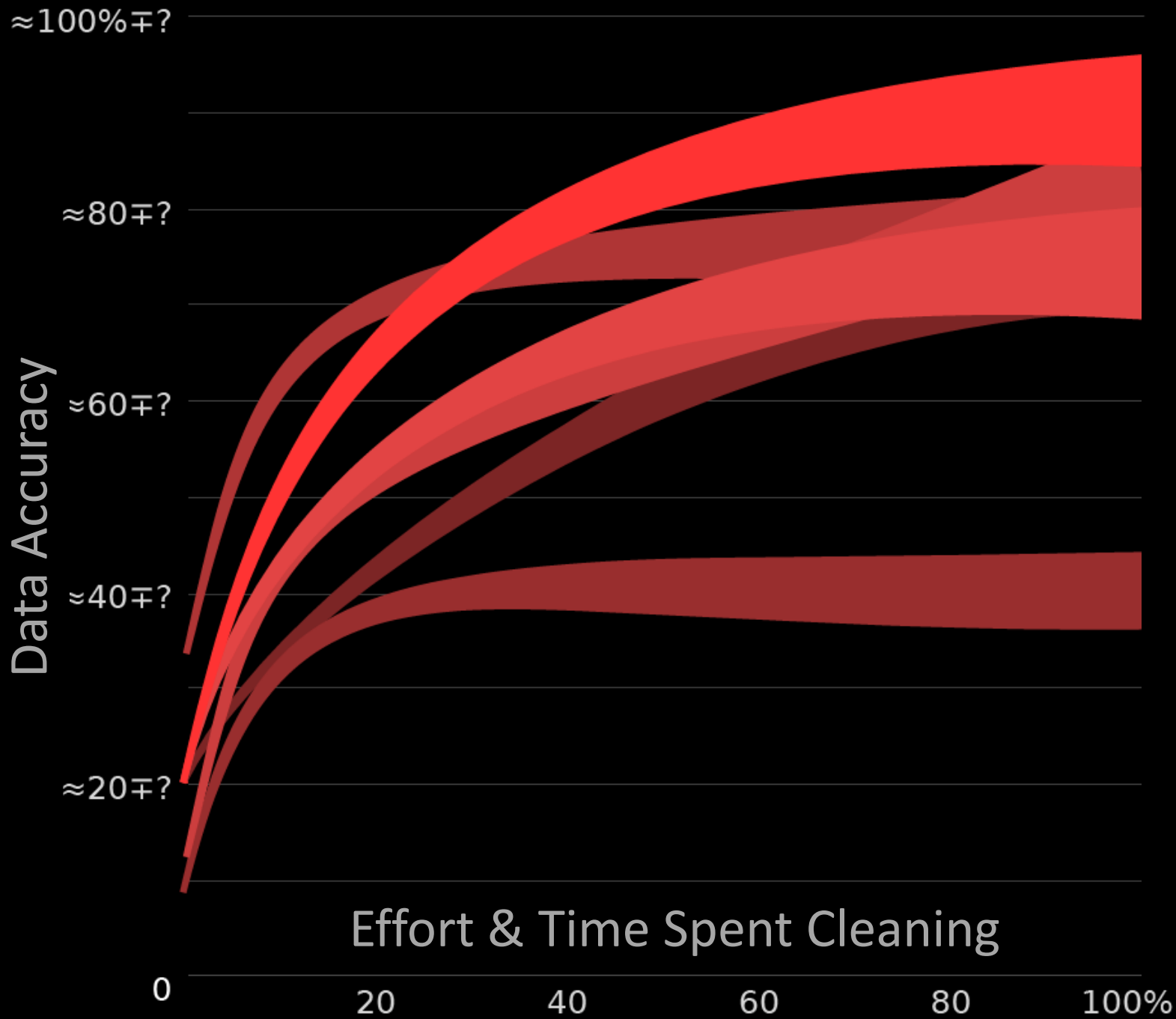


Probabilistic linkage calculates a total score for two records to determine how likely it is that both refer to the same individual. The total score is the sum of scores generated by the comparison of individually weighted fields.

Bursting the

Linkage Bubble

1. Probabilistic is better *when assumptions hold*
2. Linkage success depends on geography, ethnicity, poverty, and other health-correlated variables.
3. String comparators make a bigger difference than other tweaks to linkage methods
4. ~80% of the effort and improvement is not even in the linkage method, it's in data cleaning and preparation, but you can over-clean *and* under-clean!



What else would you like to discuss?

Name parsing

probabilistic linkage
software

Twins

using graph databases to
manage linking data

String comparators

Phonetic algorithms

request process for external
data

SSN's

Etc.!

Other data cleaning
processes, terms & issues



Patient Records Matching Overcoming the largest obstacle to health information exchange: One HIE's story

Daniel Chavez, Executive Director
San Diego Health Connect

The SDHC mission



Our Mission

To connect healthcare stakeholders to deliver quality, comprehensive information for better care.

When every individual's health information is securely available to their doctors when and where they need it:

- Doctors can provide better, more informed care.
- Duplication of tests and procedure decreases.
- Costs go down.

Participating organizations



UC San Diego
HEALTH SYSTEM

SHARP®



KAISER PERMANENTE®

Rady
Children's

Hospital
San Diego



Health
and
Human
Services



Scripps



HEALTH QUALITY PARTNERS
of Southern California

Trusted health information exchange..



Is built on
technical
interoperability

HL7
FHIR
ISO

Uses document
standards to
achieve functional
interoperability

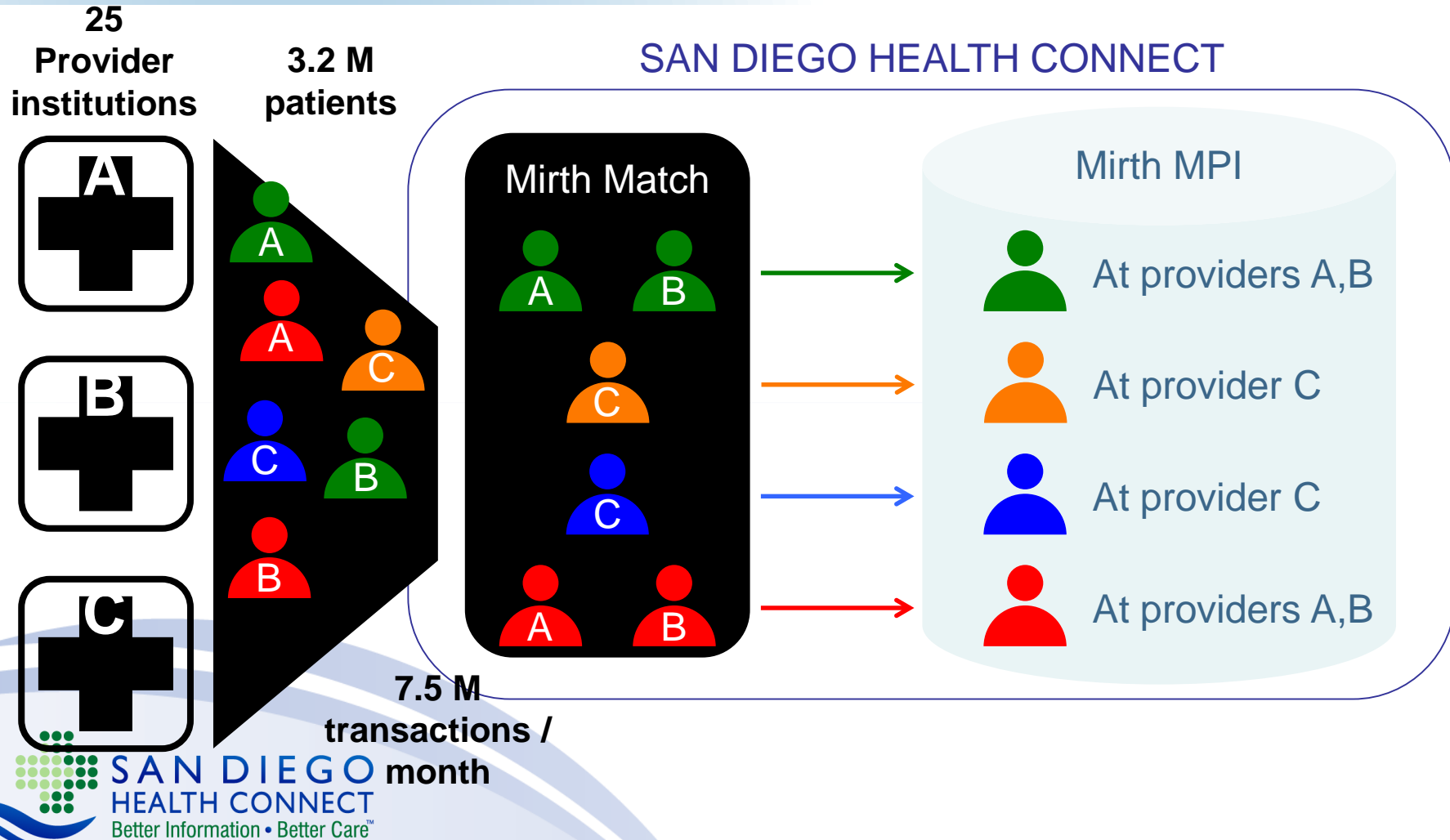
CCR
SNOMED
DICOM
LOINC

NCPDP
RxNorm
CPT
ICD-9/10

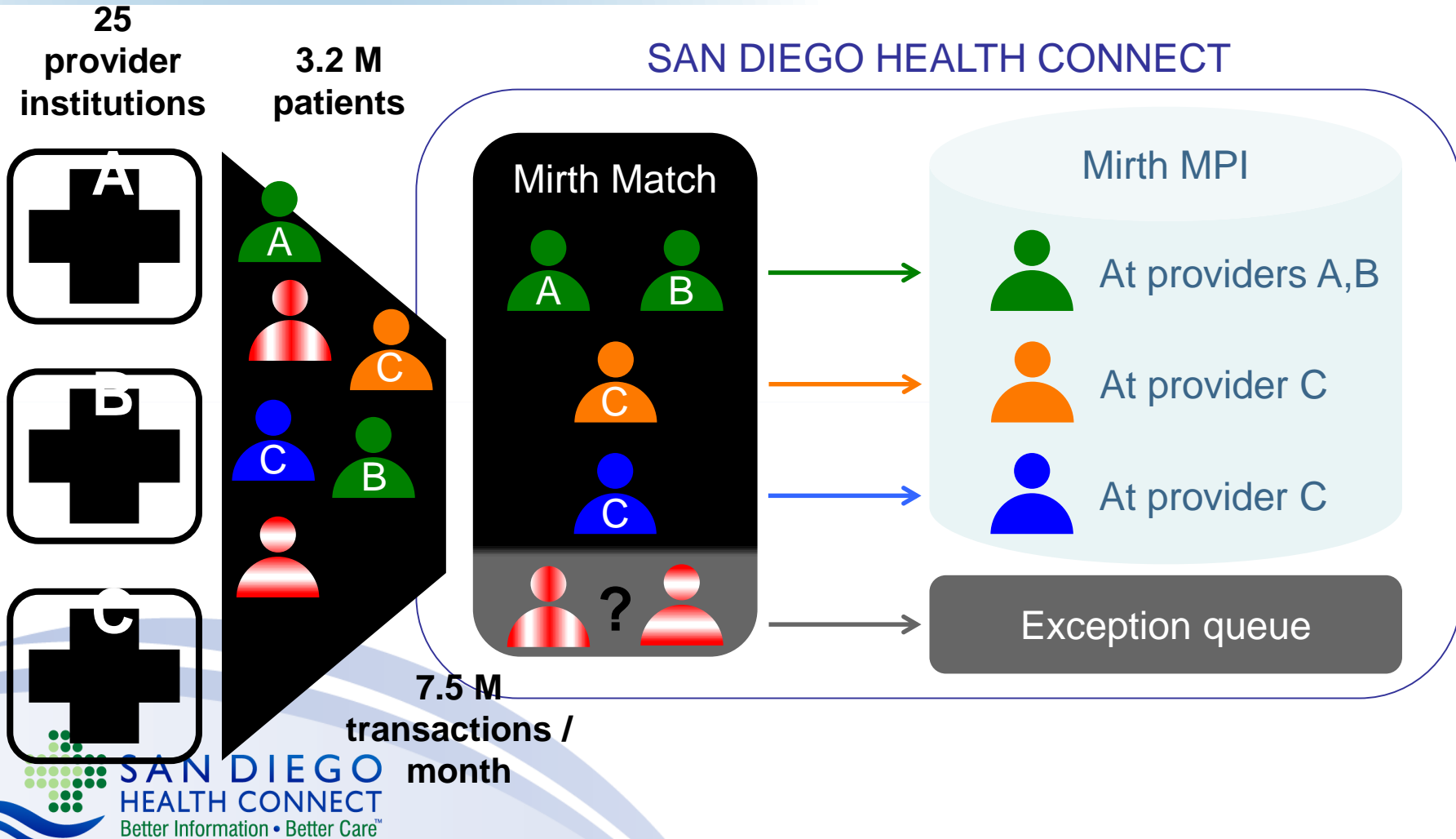
Is enabled by
semantic
interoperability

Patient Matching:
No false positives
Minimal false negatives

SDHC uses an MPI as a record locator service



When records do not match, records ended up in an “Exception queue”



Our working group decided we needed a better way to match records



41 Total members

13 Different organizations

2 Meetings per month

Referential matching is a revolutionary new way to match patient records

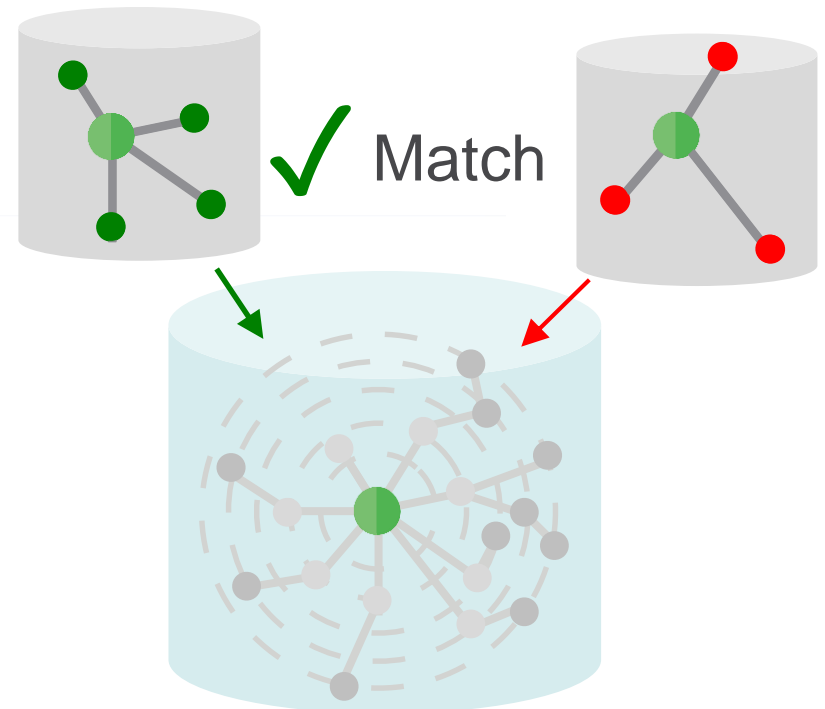


MPI matching (deterministic or probabilistic) can't see through different or bad identity data

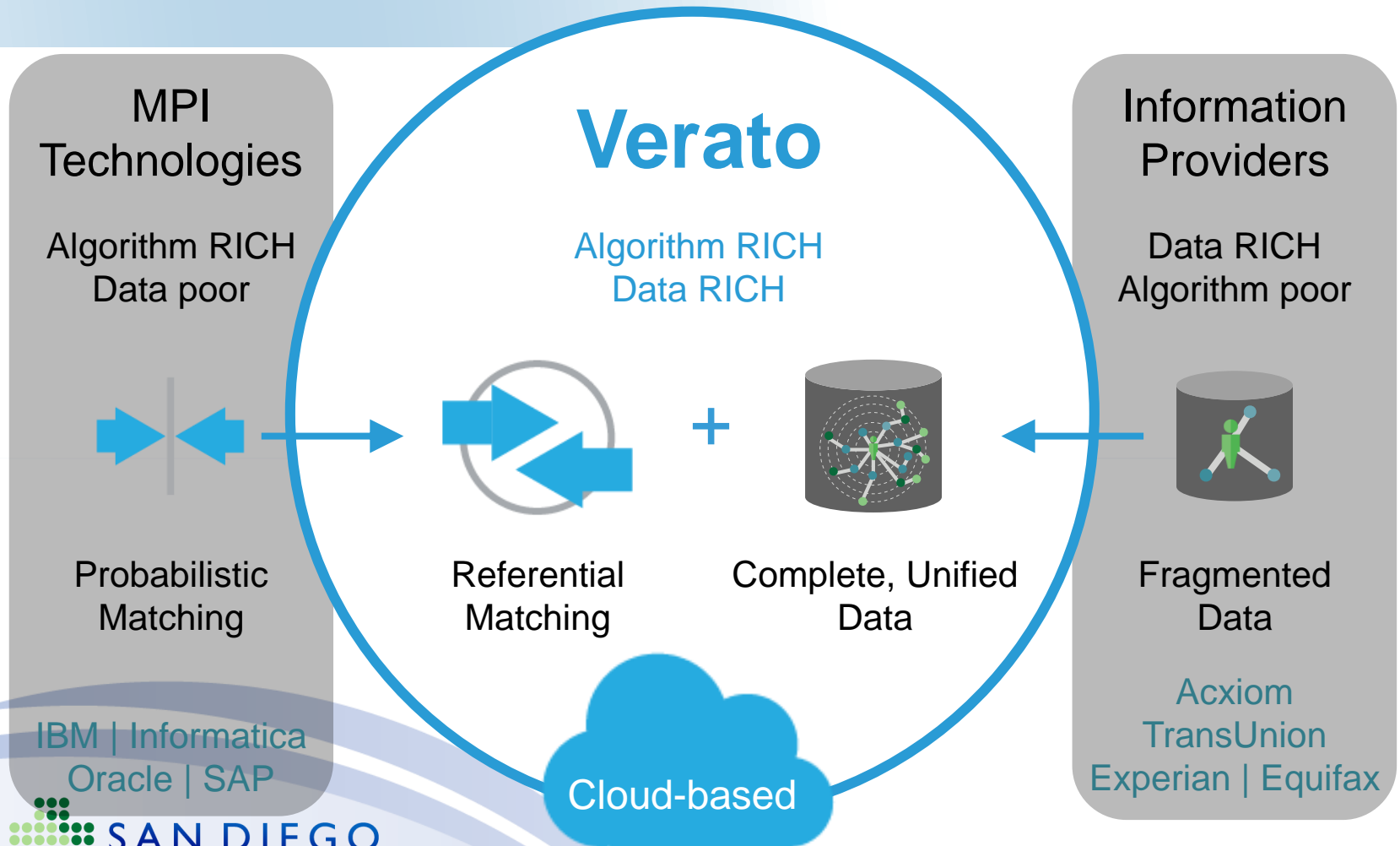


X No Match

Referential matching works despite different or bad identity data

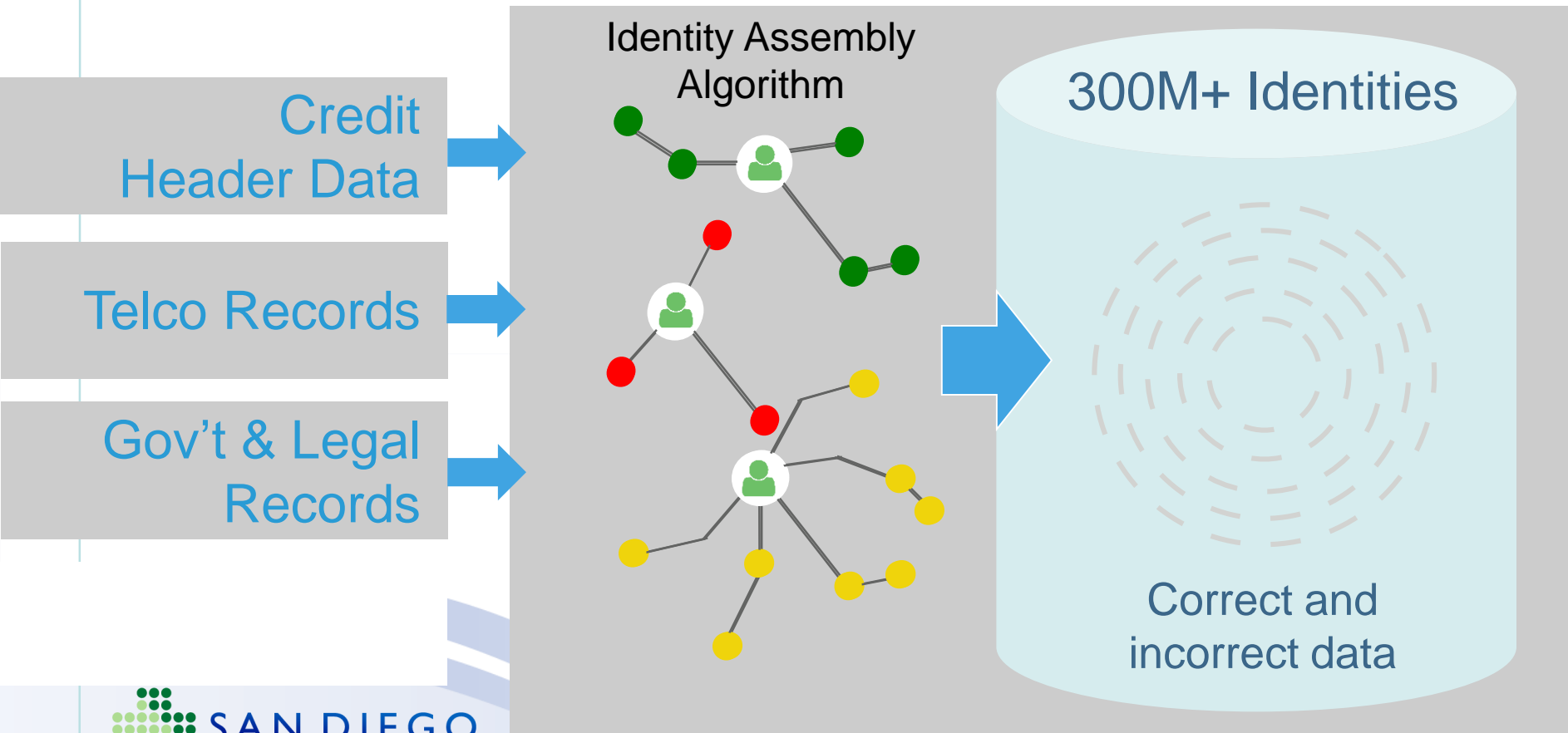


How Verato is different

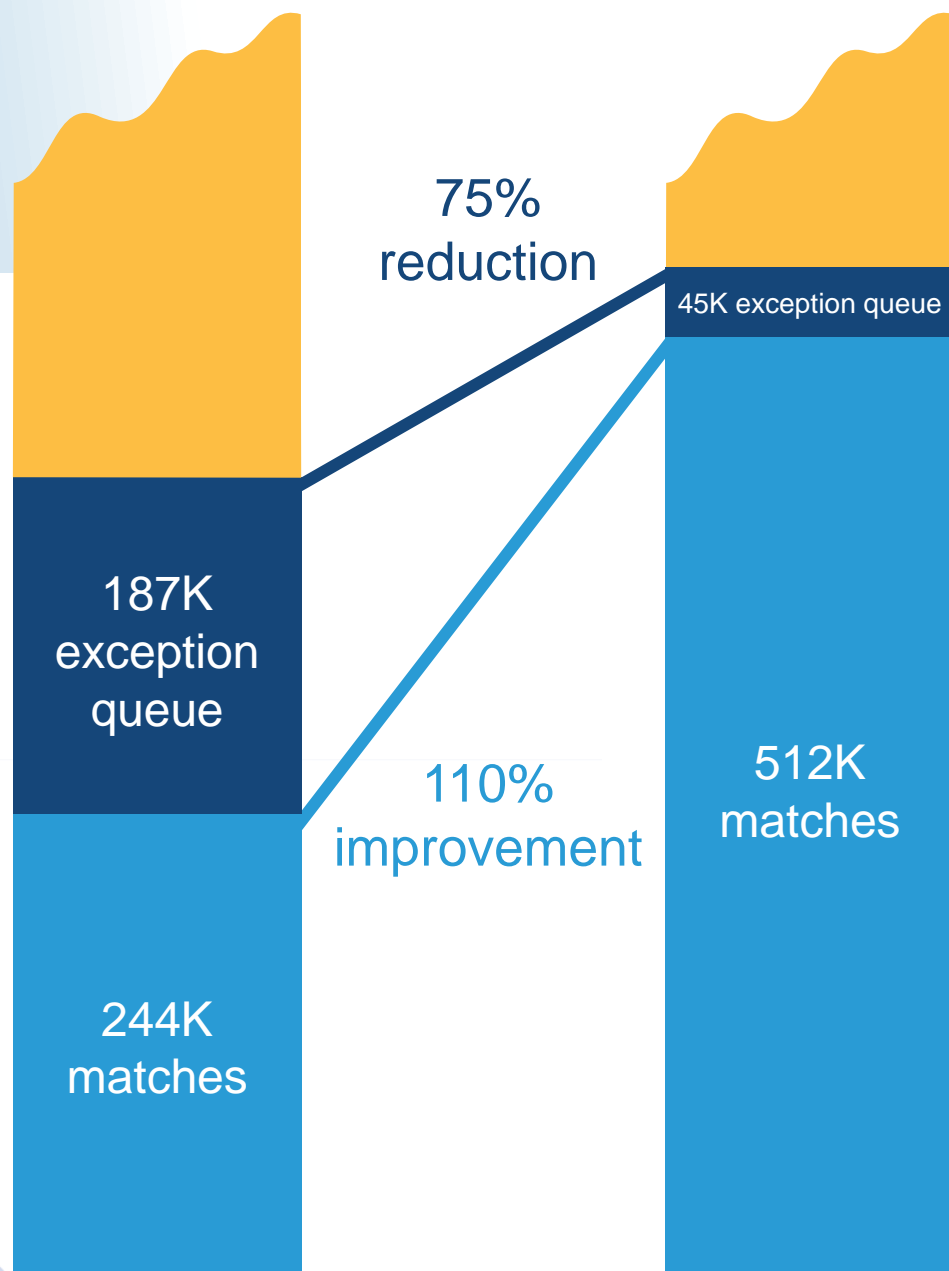


IBM | Informatica
Oracle | SAP

CARBON™ – the most comprehensive reference database of identities in the US



In total, SDHC increased the number of matches in its MPI by 110%



Futures – Improve edge case matching



1

Incorporate relationship data in ADTs

2

Pediatrics: add twins identifiers to patient data model at institutions

3

Develop twin inference algorithm for newborns to support twin analysis for adults

Futures – Accommodating varying data governance models



1

Understanding an organization's identity data governance model

2

Demonstrating proof for non-obvious matches while maximizing privacy

3

Accommodating variations in transport protocols

Futures – Connect the Community



ConnectWellSD

Connect • Collaborate • Empower



LIVE WELL
SAN DIEGO

Connecting All for Better Health & Wellness

COMMUNITY INFORMATION
EXCHANGE

@



Futures – Connect the eHealth Exchange





“Better is possible. It does not take genius. It takes diligence. It takes moral clarity. It takes ingenuity. And above all, it takes a willingness to try.”

Atul Gawande

Questions?



Presenters



Clare Tanner, PhD

Co-Director of Data Across Sectors for Health (DASH),



Melissa Moorehead

Policy Analyst and Project Manager, Michigan Public Health Institute



Stephen Singer, MCP

Senior Manager of Data Analytics, Camden Coalition of Healthcare Providers



Dan Chavez, MBA

Executive Director, San Diego Health Connect

Connect with Us!

- Sign up for news from All In at dashconnect.org
- Follow us at [@DASH_connect](https://twitter.com/DASH_connect) and [@AcademyHealth](https://twitter.com/AcademyHealth) at [#CHPHealthIT](https://twitter.com/CHPHealthIT)
- Contact information for speakers
 - Stephen Singer, stephen@camdenhealth.org
 - Dan Chavez, dchavez@sdhealthconnect.org
- [Evaluation](#)
- A resource list, slides, and recording will be available

