Data Talks

Topics we care deeply about and think you should too

AISP

Agenda

- Jessie Tenenbaum, North Carolina DHHS
- Cass Dorius, Iowa State University
- Aaron Bean, Asemio
- Amy O’Hara, Massive Data Institute

AISP

Jessie Tenenbaum

Chief Data Officer, North Carolina DHHS

AISP

Penn Social Policy & Practice
University of Pennsylvania

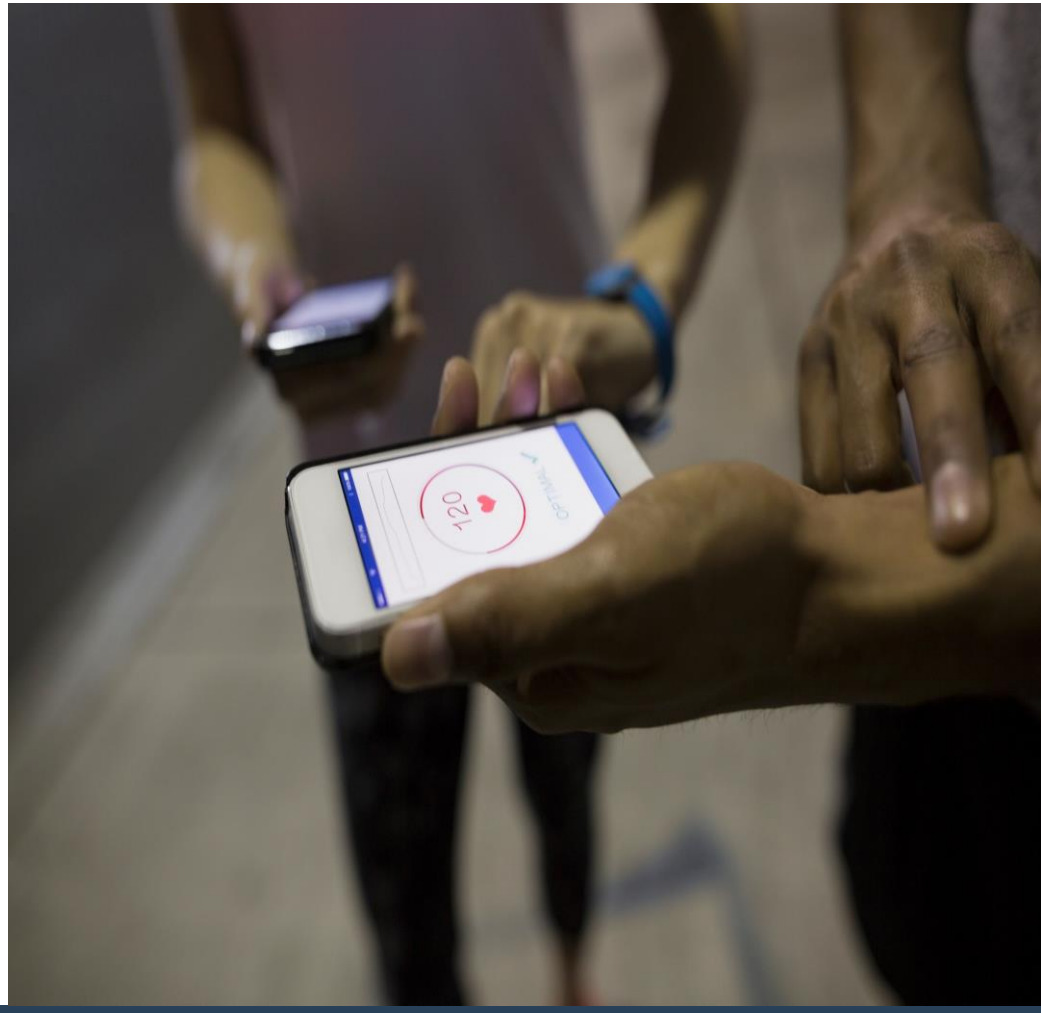
NC DHHS's IDHub

Using unique IDs to link individuals across siloed data systems.

June 2022

Jessie Tenenbaum, PhD
@jessiet1023

NC DEPARTMENT OF HEALTH AND HUMAN SERVICES

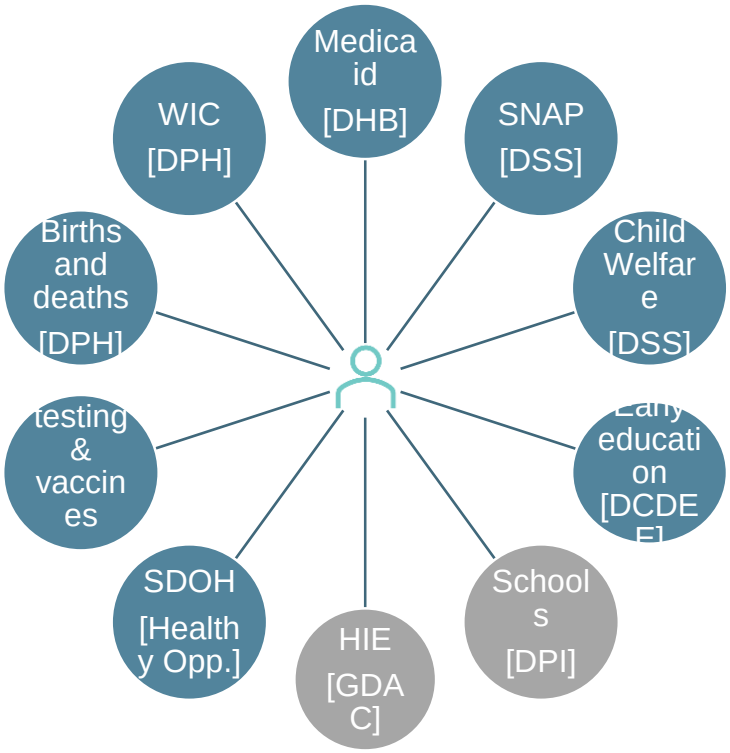


Outline

1. What problem are we trying to solve?
2. What are the options to solve it?
3. How have we gone about solving it?
4. What are lessons learned?

Motivation: Whole Person Health

- Business Goal: the ability to link data between DHHS Divisions, and other Departments, to facilitate a “Whole Person Health” view of the people we serve.
  - Real-time individual level to facilitate connecting individuals with services
  - Aggregate analysis to inform policy and enable continuous improvement
- Requires the ability to integrate data across divisional silos
- Which requires ability to link records between systems



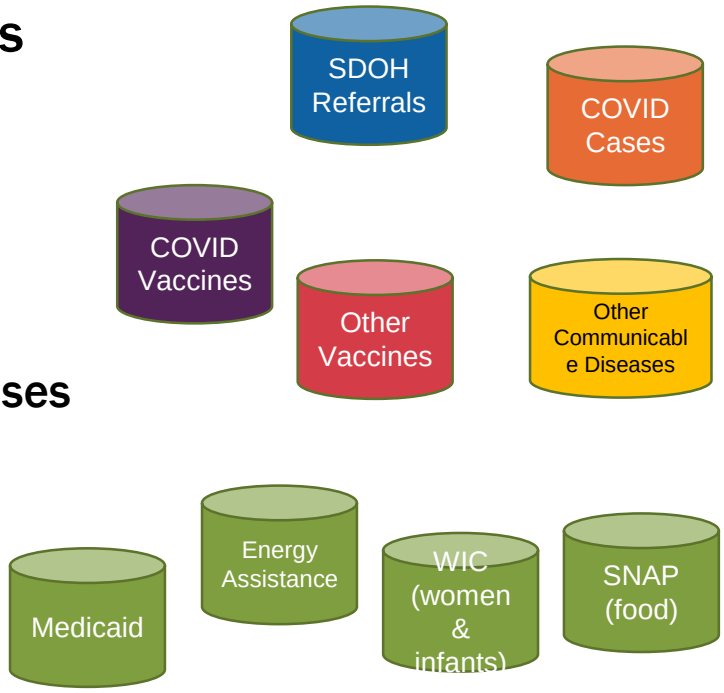
Motivating Questions (Examples)

- What % of Medicaid beneficiaries have been vaccinated?
- What is the rate of COVID tests/cases among Medicaid beneficiaries vs. the general population?
- Who has been hospitalized for COVID (vax status, demographics)?
- What % of people experiencing homelessness have been vaccinated?
- Who receives regular food assistance referrals through NCCARE360 and therefore may benefit from SNAP but is not enrolled?
- What % of kids in foster care have not been vaccinated?
- What is the relationship between early grade outcomes (e.g., third grade reading) and different early childhood conditions (e.g., early learning, health, housing, child welfare)?

Answering those questions is currently difficult at best

- Data live in silos that do not share IDs
- Probabilistic match is possible, but
  - Labor-intensive
  - Prone to error

  - Example: initial approach for post-vax cases

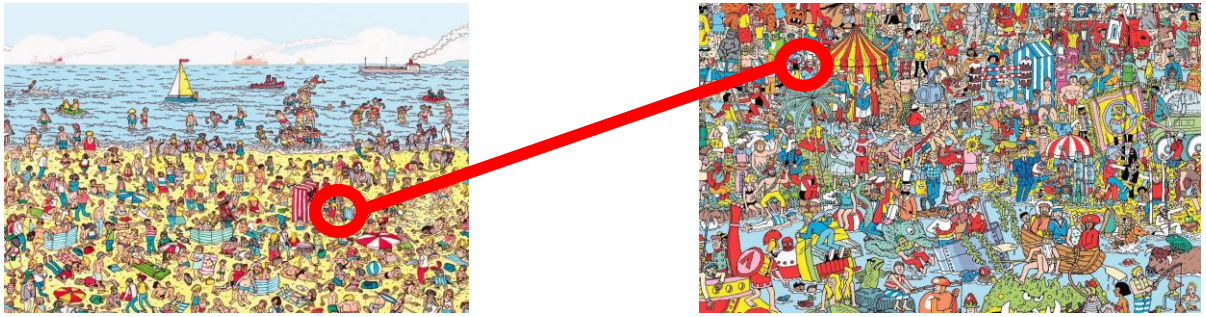


In order to achieve Whole Person Health, need 2 things:

1. **The ability to integrate data between divisional silos ("Data Integration")**

2. **The ability to identify and link the same individual from different datasets ("Entity Resolution")**

ID: 123456
Name:
Waldo

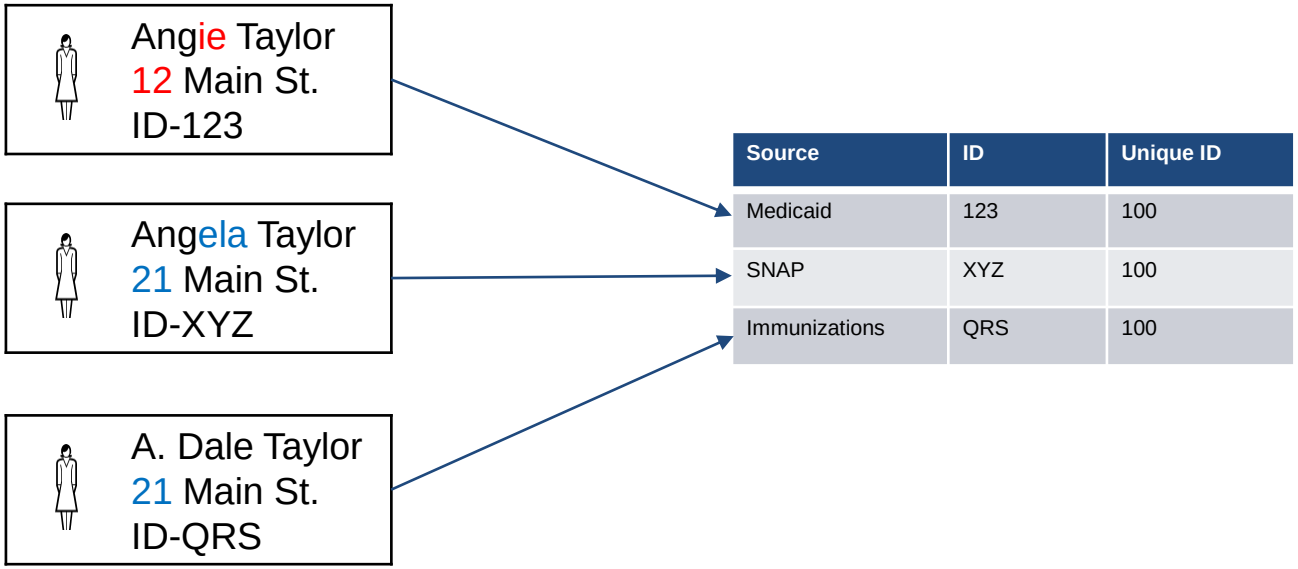


ID: 123456
Name:
Waldo

Options for “solving” entity resolution

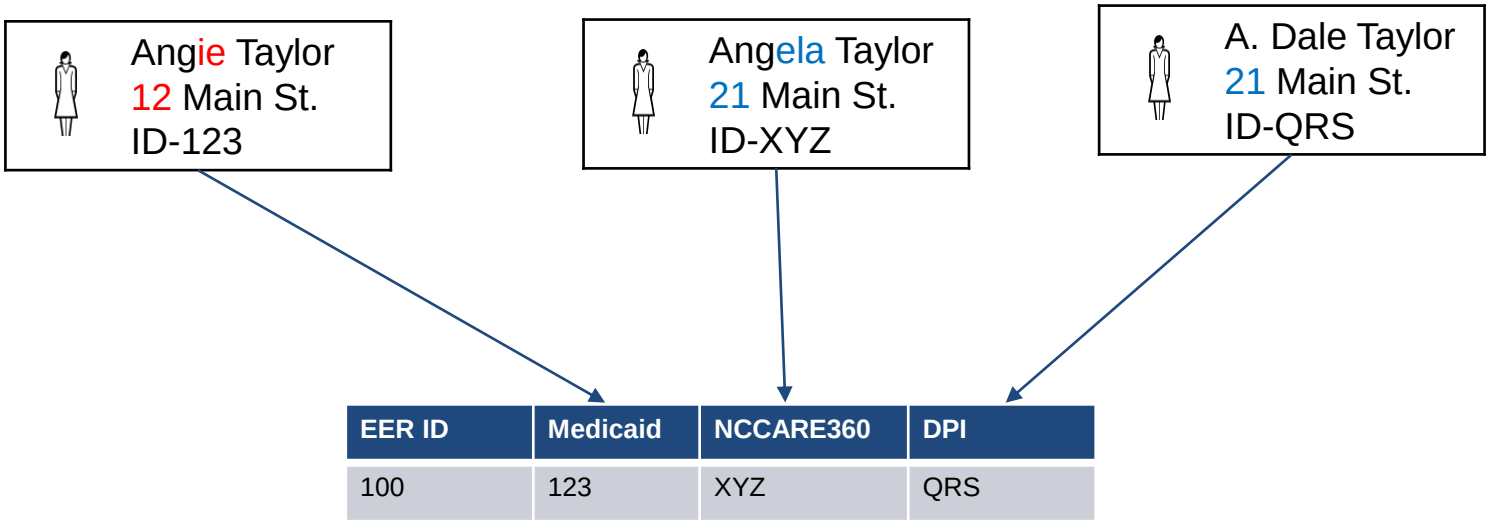
1. Each system uses its own ID, probabilistic “fuzzy match” between systems as needed
2. One ID to rule them all
3. Somewhere in between
   a. Multiple “standard” identifiers
   b. Map them to each other!
   c. Refer to that mapping for efficient data i[illegible]ation

Map each ID to a universal unique identifier…



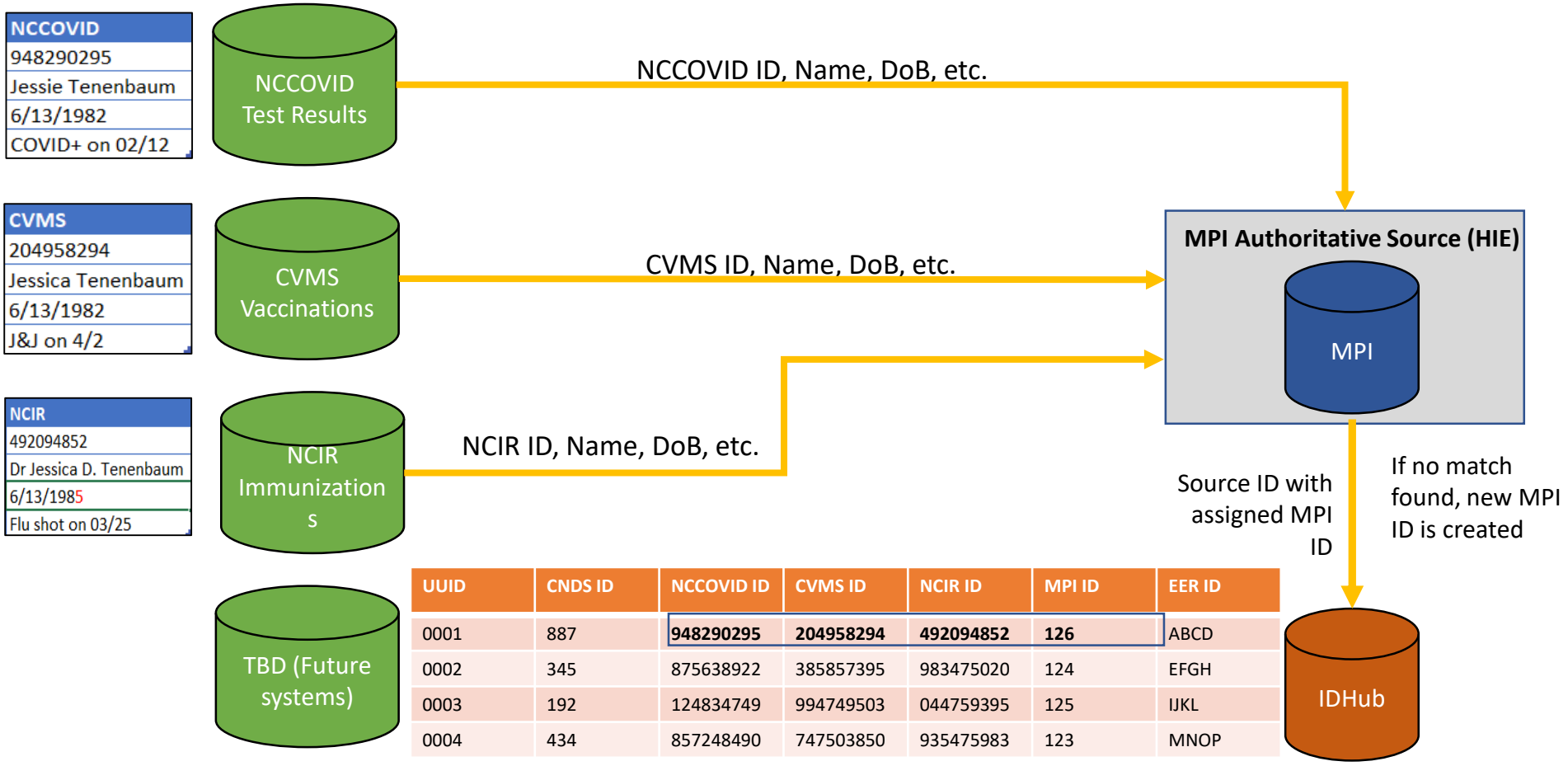
Source	ID	Unique ID
Medicaid	123	100
SNAP	XYZ	100
Immunizations	QRS	100

…enabling mapping each separate ID to the others



EER ID	Medicaid	NCCARE360	DPI
100	123	XYZ	QRS

NC DHHS Source Systems send identified attributes to ID authoritative source

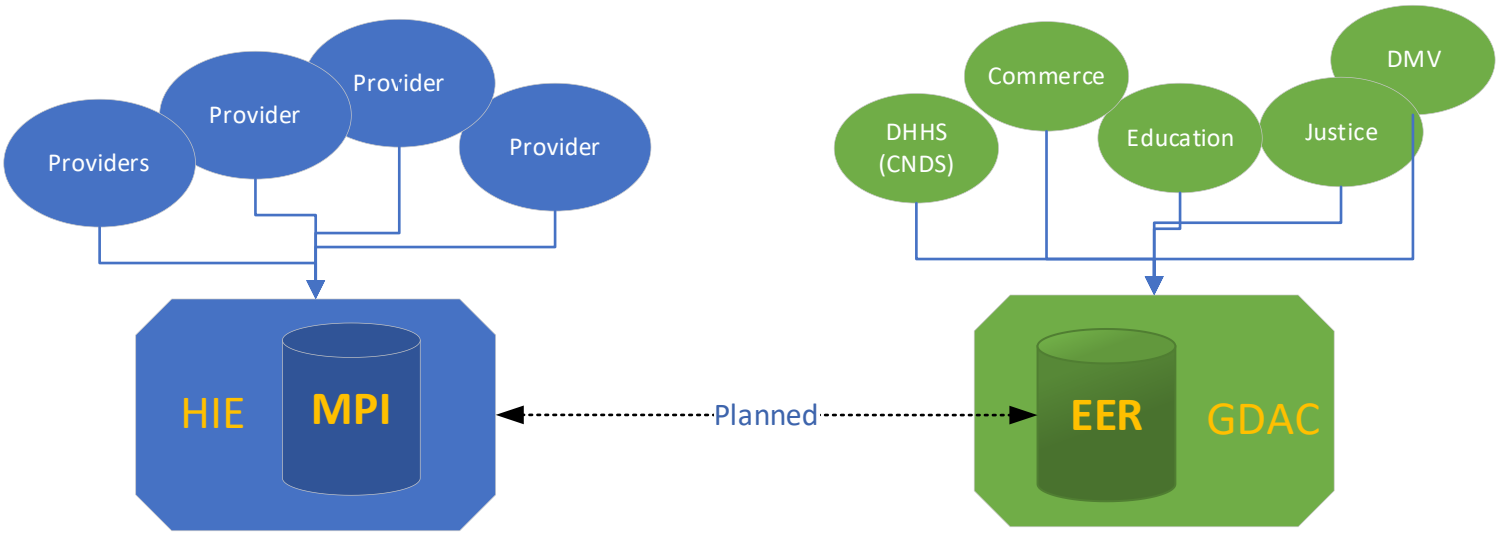


UUID	CNDS ID	NCCOVID ID	CVMS ID	NCIR ID	MPI ID	EER ID
0001	887	**948290295**	**204958294**	**492094852**	**126**	ABCD
0002	345	875638922	385857395	983475020	124	EFGH
0003	192	124834749	994749503	044759395	125	IJKL
0004	434	857248490	747503850	935475983	123	MNOP

A few things about the NC Landscape

- **State-run HIE- “NC HealthConnex”**
- **Government Data Analytics Center (GDAC)- a Division of NC’s Department of Information Technology**
  - **Created by legislation for data management and analytics**
- **NC is home to SAS**

2 “Universal” Identifiers: clinical vs. non-clinical



MPI ID (Master Patient Index)

- Used by Health Information Exchange (HIE) for **clinical data**
- Involves merging clinical records
- High stringency matching

EER ID (Enterprise Entity Resolution)

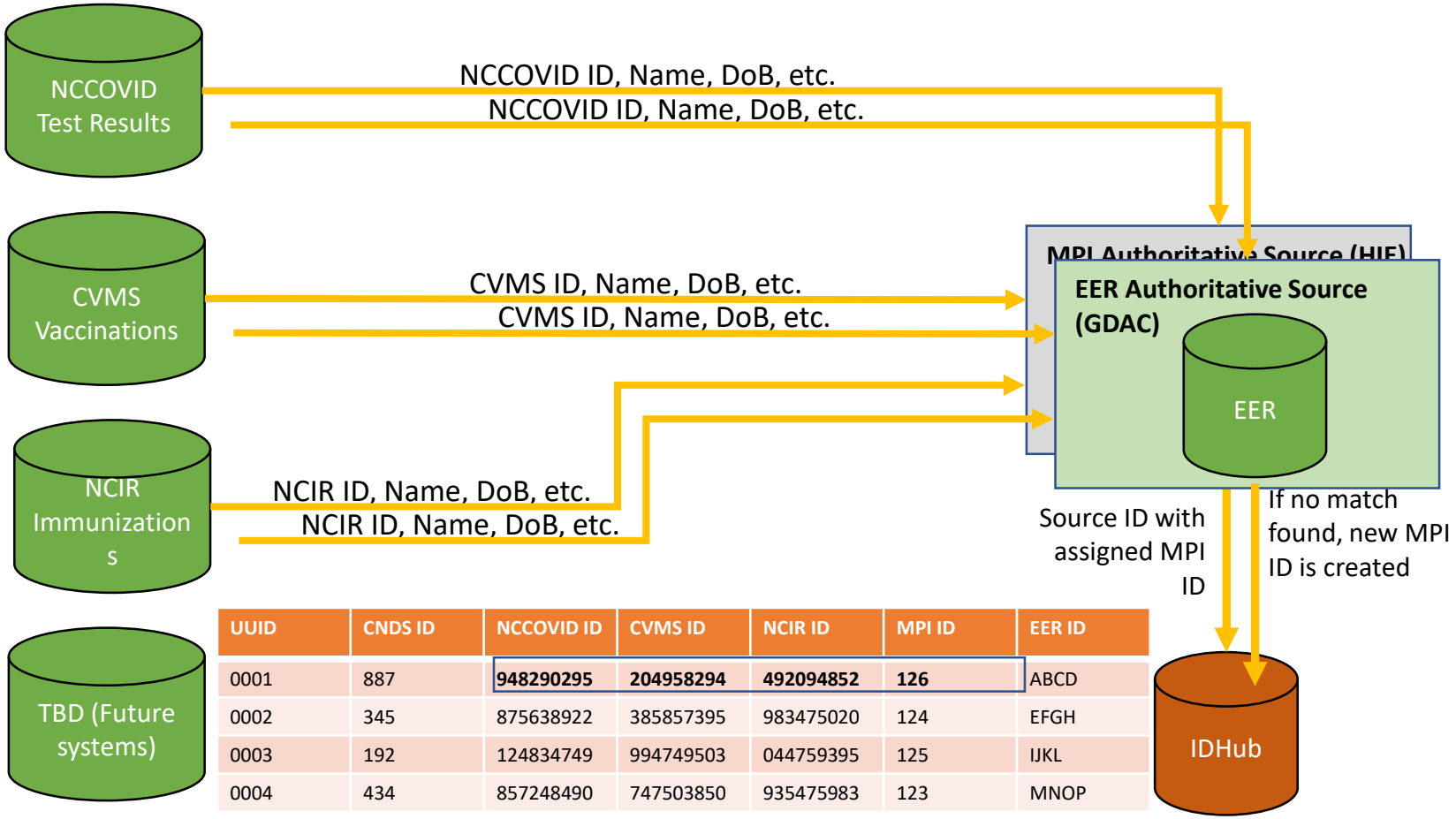
- Used by NC’s Government Data Analytics Center (GDAC) for **non-clinical data**
- Slightly lower bar for matching
- Meant for analytics

NC DHHS Source Systems send identified attributes to another ID authoritative source

NCCOVID
948290295
Jessie Tenenbaum
6/13/1982
COVID+ on 02/12

CVMS
204958294
Jessica Tenenbaum
6/13/1982
J&J on 4/2

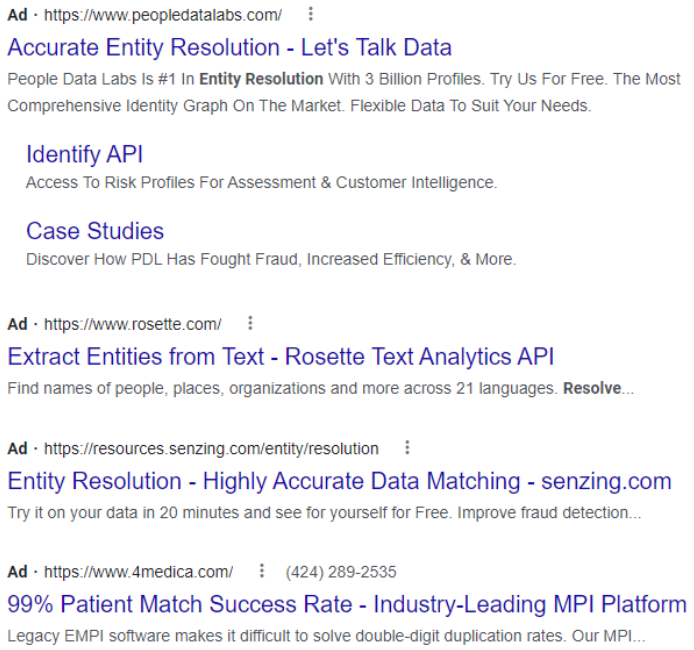
NCIR
492094852
Dr Jessica D. Tenenbaum
6/13/1985
Flu shot on 03/25



UUID	CNDS ID	NCCOVID ID	CVMS ID	NCIR ID	MPI ID	EER ID
0001	887	**948290295**	**204958294**	**492094852**	**126**	ABCD
0002	345	875638922	385857395	983475020	124	EFGH
0003	192	124834749	994749503	044759395	125	IJKL
0004	434	857248490	747503850	935475983	123	MNOP

Another option: 3rd party vendors

- **Pull from consumer data**
- **May be able to supplement with external information (for a fee)**
- **Used by some states**



MPIs can change over time

- **MPI's are frequently merged as more data are incorporated**

Jessie
Address A
MPI 111

Jessica
Address B
MPI ~~222~~ 111

Jessie
Address B
MPI 111

- **Splits (after a "false positive" match) far rarer and more challenging to handle**

Tricky things

- **Fairly obvious**
  - **Not enough data to give a confident match, e.g. Name, DoB**
  - **Twins with similar names (Zack and Jack)**
  - **Names changing over time (marriage, adoption, transgender, etc.)**
- **Learned the hard way**
  - **A provider sending the same MRN for different patients (!)**
  - **“UNK” treated as a string**
  - **Similar names and DoB’s at communal living facilities w/ same address**

Lessons Learned

- **This stuff is hard**
  - **The devil is in the details, and the edge cases**
- **It will probably take longer, and be more expensive, than you expect**
- **COVID was great for spurring action and up front $, but ongoing funding needed for operations & maintenance**
- **Concrete wins can help keep momentum**
- **Motivate business support through targeted use cases**
- **And of course- get the MOAs out of the way asap!**

Thanks!

Jessie.Tenenbaum@dhhs.nc.gov

@jessiet1023

Cass Dorius

Associate Professor, Iowa State University

AISP

Penn Social Policy & Practice
University of Pennsylvania

Driving Equity with Integrated Data: A (Mostly) non-Technical Primer

June 22, 2022

Cass Dorius, Heather Rouse & Todd Abraham

i2d2 Iowa's Integrated Data System for Decision-Making

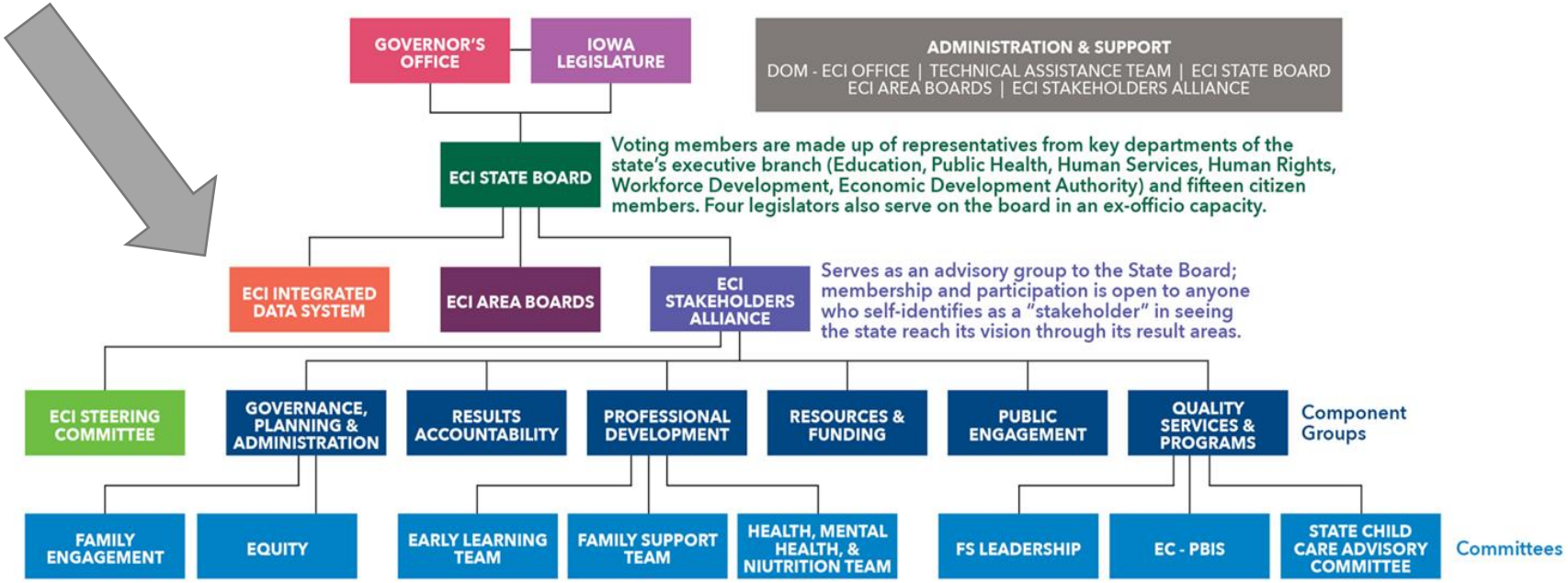
I2D2 Vision

Our system of early childhood policies and programs across the state will be informed by rigorous analysis of timely, comprehensive, and integrated data from health, human services, and education systems. Implications of policy and program analysis will be considered in collaboration to direct services and resource allocation.

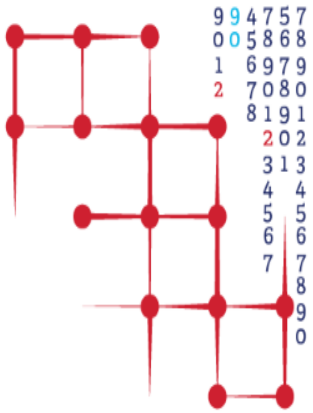
i2d2 Iowa's Integrated Data System for Decision-Making

Yay! Oh, no!

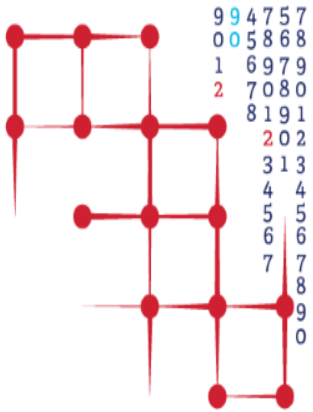
HOW THE CURRENT SYSTEM IS STRUCTURED



Systems Thinking



- IDSs are part of systems with long histories of power inequities
- Systems are hard to change
- Real people (data points) are also part of the system
- Our goal is to capture the experience of BIPOC clients in more authentic and accurate ways without further stigmatization
- A system-level problem requires a system-level response

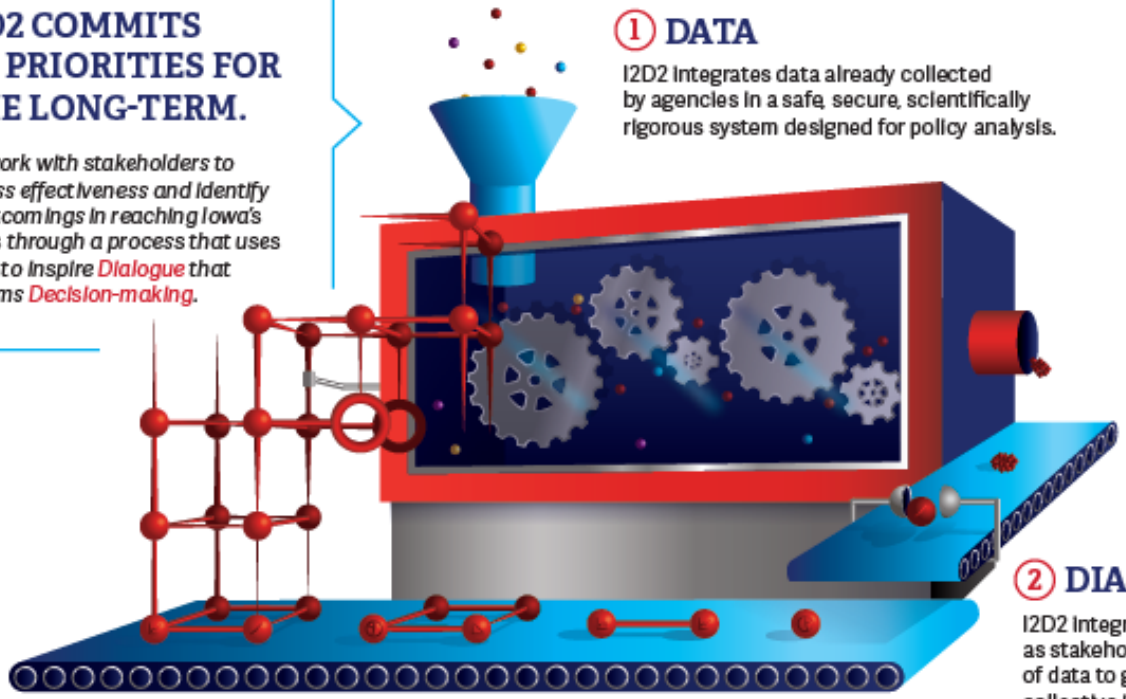


What do you see?

I see White people.

I2D2 COMMITS TO PRIORITIES FOR THE LONG-TERM.

We work with stakeholders to assess effectiveness and identify shortcomings in reaching Iowa's goals through a process that uses Data to inspire Dialogue that informs Decision-making.



1 DATA

I2D2 integrates data already collected by agencies in a safe, secure, scientifically rigorous system designed for policy analysis.

2 DIALOGUE

I2D2 integrates people as stakeholder stewards of data to gather collective insight and translate findings into actionable intelligence.

3 DECISION-MAKING

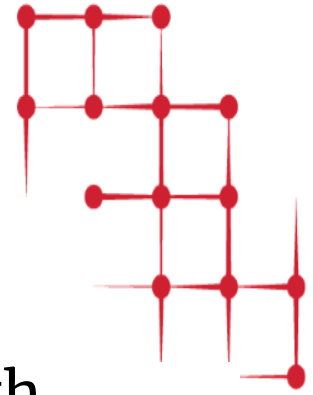
I2D2 integrates data insights with executive leader and program manager decision-making to advance a statewide culture of evidence-based services to improve outcomes.

Learn more at I2D2.iastate.edu

Equity in I2D2

Ongoing Conversations

One of the most powerful ways to promote equity is to keep the conversation going

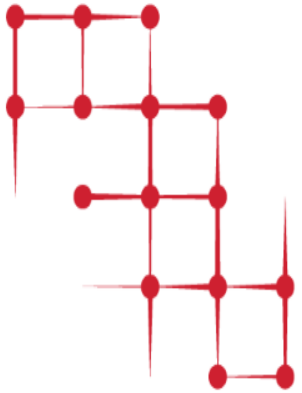


- Early conversations with agency directors and ECI board of directors about need for equity
- ‘Road show’ of data equity concerns with state partners (AISP paper)
- Reading about and sharing positive practices used by others
- Ongoing discussions about how to integrate equity throughout our processes- this requires iteration
- Thinking about who is ‘in’ and who is ‘out’ of the data system

Equity in I2D2

Expanded Data Preparation

Create a solid foundation for equitable integrations by retaining as many voices (cases) as possible



- Re-focus data preparation to be more inclusive *before* we start integrating
- Create space for diversity-look for, clean, and categorize diverse naming conventions such as hyphenated, compound names, multiple names, and partial names
- Human verification makes sure our decision rules "work"

Equity in I2D2

Multi-step Process

Think about the step-wise process needed to match complex names and non-standard dates (regardless of your algorithm choices)

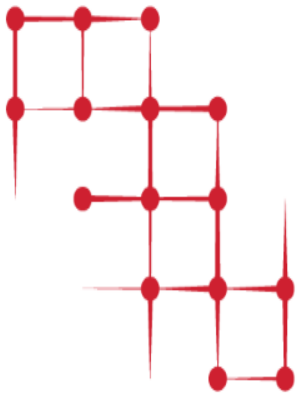
- Missing cases & incorrect matching greatly impact rural data accuracy
- Missing data also disproportionately silences the voices of immigrants, refugees, and BIPOC
- We added steps to capture as many voices as possible by linking (1) complex naming conventions; matching first, middle, and last names, and (2) US and Non-US dates such as 6/7/2022 versus 7/6/2022

Equity in I2D2

Human-led Algorithms

"There is no such thing as a race-neutral algorithm"

Avoid 'black box' algorithms. Use multiple algorithms to address linking needs. Order of operations matters! Human verification is critical.



Current Process – Multiple Passes & Order of Operations

1. Original Records
2. Collapsed Records
   (delaCruz = delaCruz ≠ delaSalle)
3. Selectively Lowered Cutoff Criteria
   (Kathryn ≈ Katie; Ignacio ≈ Nacho)
4. Tighter Cutoff Criteria on Multiple Fields
5. Standardized Decision Rules
6. Human Verification/Rejection

We would love to hear from you:

i2d2.iastate.edu

i2d2

Iowa's Integrated Data System for Decision-Making

Aaron Bean

Managing Partner, Asemio

AISP

Penn Social Policy & Practice
University of Pennsylvania

Data & Technology Ecosystems

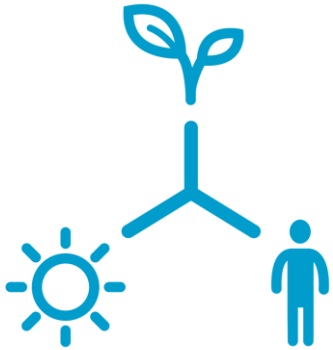
Measuring advancement and innovation

Aaron Bean, *Asemio*

asemio AISP

What are Data & Technology Ecosystems?

ec·o·sys·tem
ˈēkōˌsistəm/
noun



ECOLOGY

1. a biological community of interacting organisms and their physical environment.
    - (in general use) a complex network or interconnected system.
    - "Silicon Valley's entrepreneurial ecosystem"

The value of measuring D&T ecosystem innovation.

- Collective Impact initiatives
- Value-based care
- Pay for success contracts
- Actionable insights
- Evidence-based policymaking

- Integrated data system implementation
- Data warehouse project
- Data modernization efforts
- Community Information Exchange
- Dashboard | Portal | Longitudinal Evaluation System

Challenge

Why the lack of investment measurement models?

Challenge

Why the lack of investment measurement models?

Public Interest Data and Technology Ecosystem Development as a wicked problem

Wicked Problem Characteristics

1. There is no definitive formulation of a wicked problem
2. Wicked problems have no stopping rule
3. Solutions to wicked problems are not true-or-false, but good-or-bad
4. There is no immediate and no ultimate test of a solution to a wicked problem
5. Every solution to a wicked problem is a "one-shot operation"; because there is no opportunity to learn by trial-and-error, every attempt counts significantly
6. Wicked problems do not have an enumerable (or an exhaustively describable) set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into the plan
7. Every wicked problem is essentially unique
8. Every wicked problem can be considered to be a symptom of another problem
9. The existence of a discrepancy representing a wicked problem can be explained in numerous ways. The choice of explanation determines the nature of the problem's resolution
10. The planner has no right to be wrong

–Rittel, Horst. "Dilemmas in a General Theory of Planning." Policy Sciences, 1973: 155-169.

Taming Data & Technology Ecosystem Development

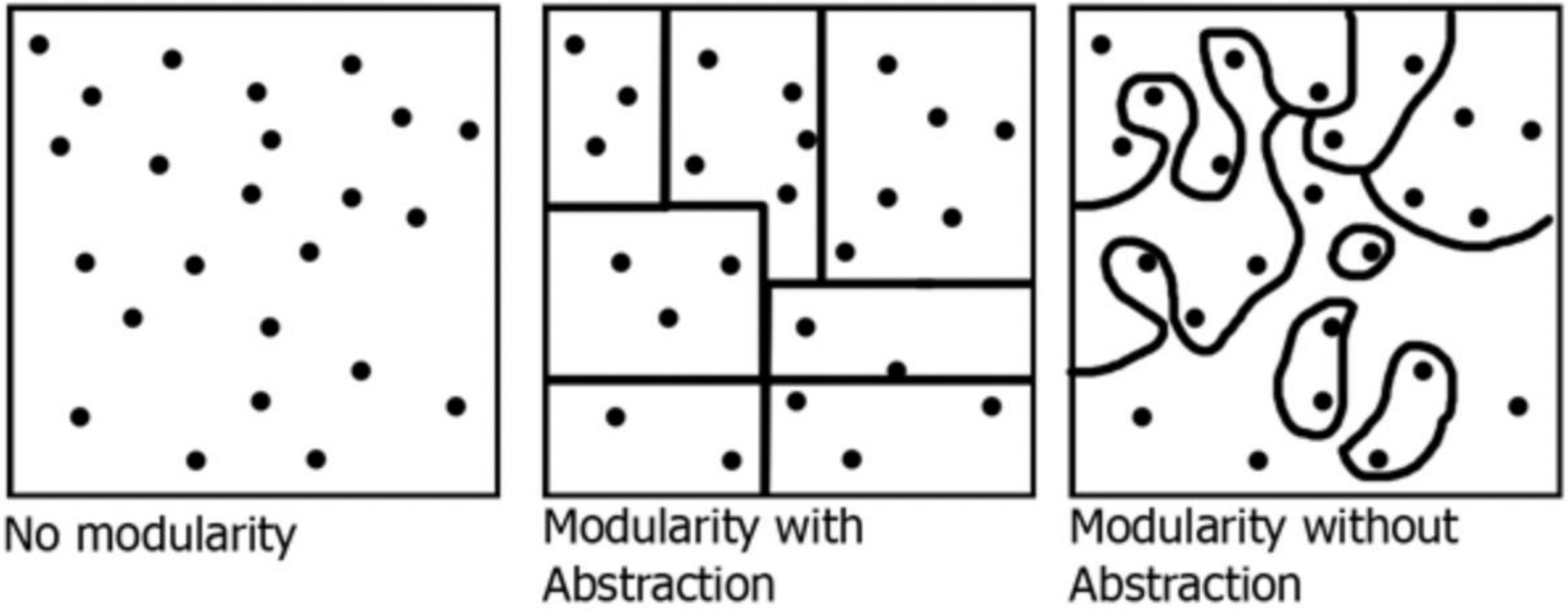
We find that the emergence of a ‘modular governance architecture’ has helped diverse and heterogeneous actors turn sustainability from an ambiguous concept into a concrete set of semi-independent practices, while mitigating governance complexity.

Stephan Manning, Juliane Reinecke,
A modular governance architecture in-the-making: How transnational standard-setters govern sustainability transitions,
Research Policy, Volume 45, Issue 3, 2016, Pages 618-633, ISSN 0048-7333,
https://doi.org/10.1016/j.respol.2015.11.007.

The modularity principle is the most basic principle in engineering. It states:

```
Systems should be built from cohesive, loosely
coupled components (modules).
```

http://www.cs.sjsu.edu/~pearce/modules/lectures/ood/principles/Modularity.htm

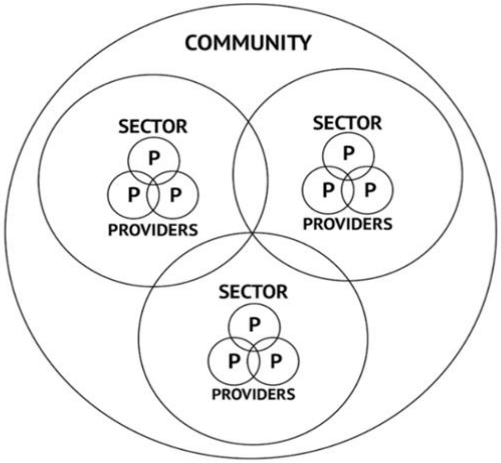


http://web.cs.ucla.edu/classes/winter12/cs111/scribe/3a/

Data Ecosystem Framework

POINT OF VIEW

LEVEL OF WORK



AREA OF ADVANCEMENT

CULTURE

Speaks to the community data vision, resource commitment, and level of cooperation and trust.

Community Vision

Resource Commitment

Cooperation and Trust

GOVERNANCE

Speaks to regulatory influence, community policy, and overall sustainability model.

Privacy and Security

Policy

Sustainability Plan

INFRASTRUCTURE

Speaks to knowledge management, level of fidelity to systems change models, and technical capabilities.

Knowledge Management

Methodological Framework

Technology Infrastructure

Example Questions to Measure Innovation

Improved speed of data sharing

- "Does this solution decrease the **time to answer (TTA)?"**
- What is the time from project start to report delivery? How does this compare to a baseline?

Improved ease of use

- "Is this solution **easier** to use?"
- What are the indicators from stakeholders that the overall experience is better?

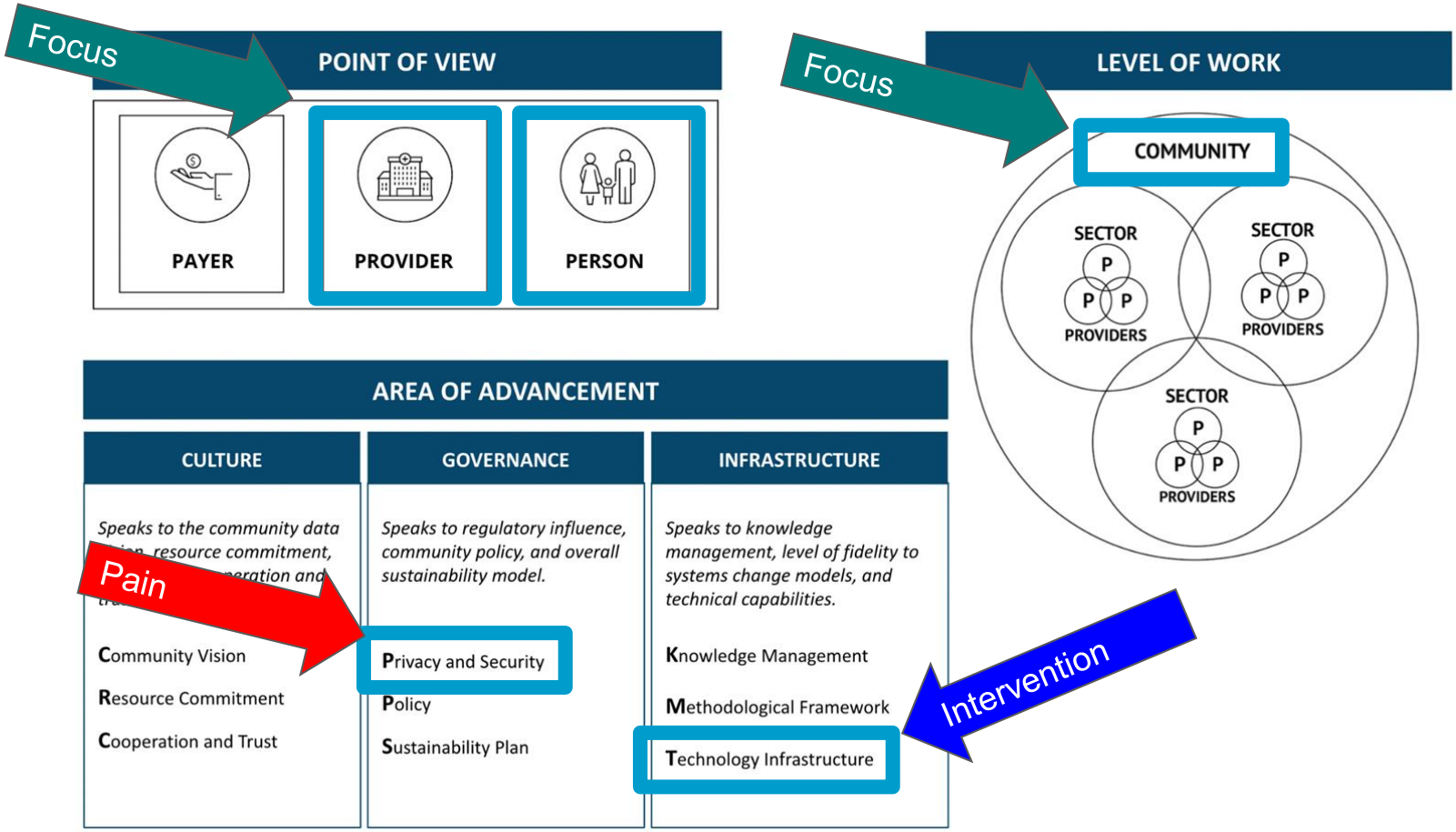
Improved lateral interoperability

- "Does this solution facilitate **lateral interoperability**?"
- Can data be combined and analyzed across "same-level" sources (eg., all community-based organizations or all state agencies) and between "same-level" analytics systems?

Improved vertical interoperability

- "Does this solution facilitate **vertical interoperability**?"
- Can data be combined across "different-level" (i.e., local, state, federal) data sources and analyzed between "different-level" analytics systems?

Data Ecosystem Framework



PET CASE STUDY

GEORGE KAISER FAMILY FOUNDATION
A supporting organization of Tulsa Community Foundation

BIRTH THROUGH EIGHT STRATEGY FOR TULSA

blue meridian partners

Community Level **Success** with Data Sharing

"How do we make data sharing **easier** and **more accessible** in Tulsa?"

THE RESULTS

Data from a **12-year period** accessed from **32** programs
Participation from **8** organizations
Insights to **5** community questions
½ the time invested when compared with traditional methods

… all achieved in ***2 months' time***

Thank you!

asemio AISP

Amy O'Hara

Research Professor, Massive Data Institute

(No slides)

AISP

Penn Social Policy & Practice
University of Pennsylvania

Thank You