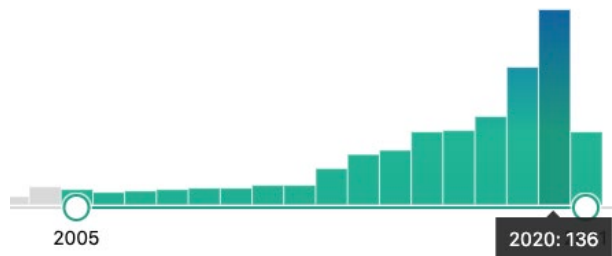


# The Indiana Addictions Data Commons (IADC)

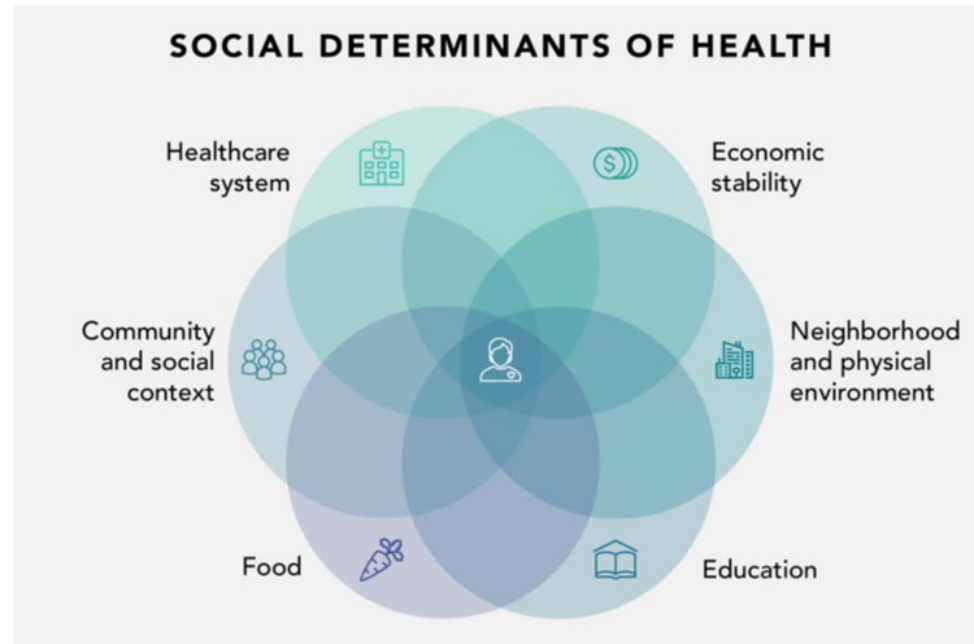
A Growing Source of Integrated  
Clinical and Social Data

# Problem

- Growing demand to integrate clinical, social, and other exposome data
- This demand is highly relevant for those working in the addictions space

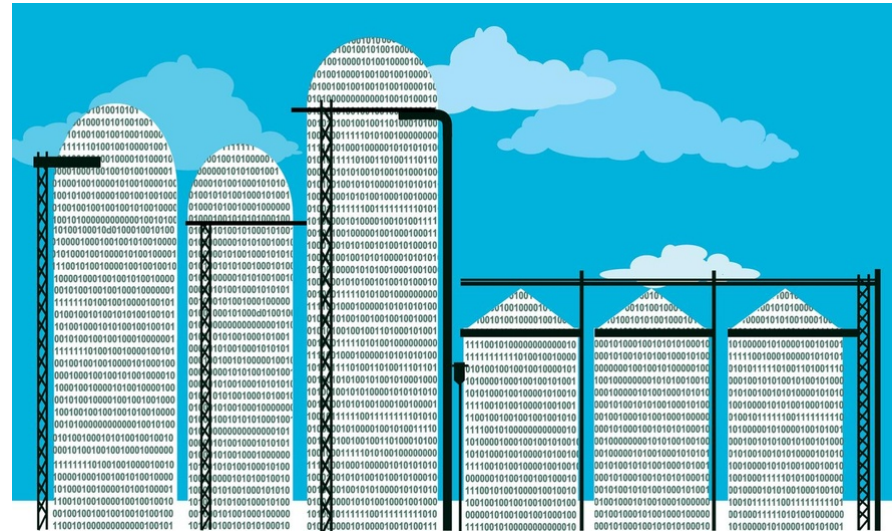


Pubmed search: social determinants of health +  
addiction



# Problem

- **Good News:** Many clinical and social determinants of health data resources exist across state and local organizations
- **Bad News:** Data is siloed; no unified process exists to access the data OR integrate the data

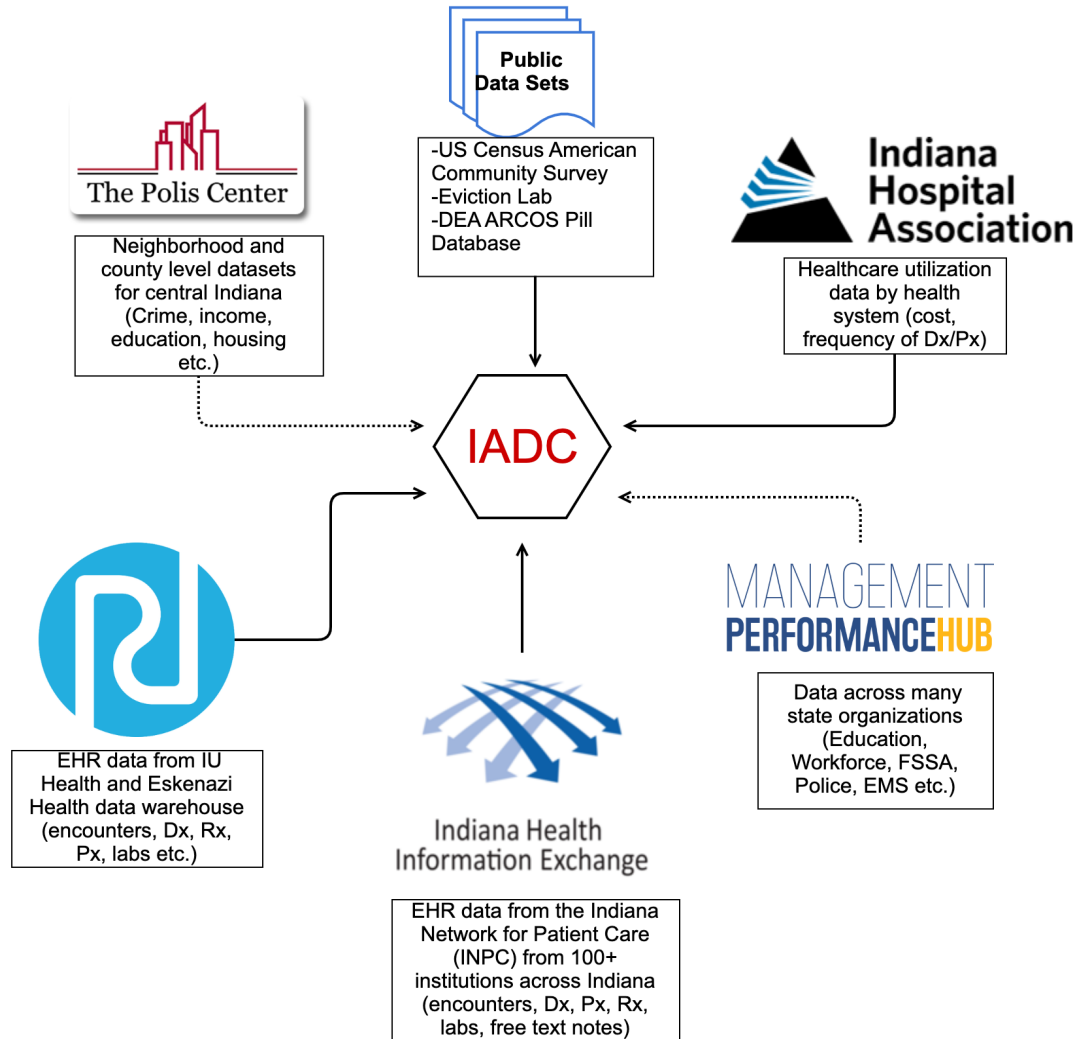


# Indiana Addictions Data Commons

- **Vision**
  - Foster improvements in health innovation and discovery by creating a state-of-the-art common data resource that will serve as the core data and information hub for addressing the addiction crisis.
- **Mission**
  - Address the addiction crisis by making robust, comprehensive, and commonly accessible data available to physicians, researchers, public health, and policy-makers in order to appropriately characterize, address, and monitor the crisis.

# The IADC Network

- Data accessible through the IADC comes in a variety of forms
  - Link data between clinical and social domains
  - Person level as well as geographic
  - Access to data sources can vary depending on request
- Access to the IADC is mediated by a data concierge and honest data broker
  - Facilitates governance process
  - Develops data request
  - Hands off integrated data set
- Future direction is focused on more self service capabilities



# Data Inventory and Prioritization

- Data inventory and prioritization workstream
  - Lit review was performed to assess the current efforts across the country to integrate social determinants data with clinical data
  - Data needs for each publication were categorized in 1 of 7 different groups
  - In-person interviews were also conducted with many IU Addictions Grand Challenge awardees and their data needs were categorized in a similar grouping

| SOCIAL DETERMINANTS CATEGORY                 | EXAMPLE MEASURES  |   |
|--|---|---|
| Socioeconomic status and material conditions | <ul style="list-style-type: none"> <li>• Income</li> <li>• Poverty</li> <li>• Access to food</li> <li>• Employment</li> </ul> | <ul style="list-style-type: none"> <li>• Living conditions</li> <li>• Race/ethnicity</li> <li>• Gender</li> <li>• Insurance status</li> </ul> |
| Behaviors                                    | <ul style="list-style-type: none"> <li>• Smoking and tobacco use</li> <li>• Diet</li> <li>• Illicit substance use</li> </ul>  | <ul style="list-style-type: none"> <li>• Alcohol use</li> <li>• Medication adherence</li> <li>• Physical activity</li> </ul>                  |
| Built environment                            | <ul style="list-style-type: none"> <li>• Transportation</li> <li>• Sidewalks</li> </ul>                                       | <ul style="list-style-type: none"> <li>• Walkability</li> <li>• Buildings</li> </ul>  |
| Natural environment                          | <ul style="list-style-type: none"> <li>• Air quality</li> <li>• Pollution</li> </ul>  | <ul style="list-style-type: none"> <li>• Climate</li> <li>• Greenspace</li> </ul>   |
| Public policies                              | <ul style="list-style-type: none"> <li>• Health policies</li> <li>• Social policies</li> </ul>                                | <ul style="list-style-type: none"> <li>• Laws</li> <li>• Regulations</li> </ul>   |
| Health services and conditions               | <ul style="list-style-type: none"> <li>• Access to health care</li> <li>• Utilization</li> </ul>                              | <ul style="list-style-type: none"> <li>• Health literacy</li> <li>• Disease prevalence</li> </ul>   |
| Social circumstances                         | <ul style="list-style-type: none"> <li>• Family</li> <li>• Social support</li> <li>• Caregivers</li> </ul>                    | <ul style="list-style-type: none"> <li>• Marital status</li> <li>• Civic participation</li> <li>• Community stigma</li> </ul>                 |






# Data Inventory and Prioritization

## Results of data integration literature review

Table 1. Characteristics of literature on social determinants of health used in combination with clinical patient-level data

|                             | Total | Social determinant categories              |           |                   |                     |                 |                              |                      |
|-----------------------------|-------|--|-----------|-------------------|---------------------|-----------------|------------------------------|----------------------|
|                             |       | Socioeconomic status & material conditions | Behaviors | Built environment | Natural environment | Public policies | Health services & conditions | Social circumstances |
|                             | n=178 | n=161                                      | n=20      | n=25              | n=8                 | n=2             | n=20                         | n=45                 |
| <b>Clinical data source</b> |       |  |           |                   |                     |                 |                              |                      |
| EHR                         | 63%   | 62%  | 90%       | 64%               | 75%                 | 50%             | 85%                          | 78%                  |
| Registry                    | 19%   | 20%  | 10%       | 16%               | 25%                 | 0%              | 10%                          | 16%                  |
| Claims / Discharge          | 20%   | 11%  | 0%        | 8%                | 0%                  | 0%              | 5%                           | 0%                   |
| Various                     | 12%   | 7%   | 0%        | 12%               | 0%                  | 50%             | 0%                           | 7%                   |
| <b>Outcomes</b>             |       |  |           |                   |                     |                 |                              |                      |
| Utilization                 | 34%   | 37%  | 10%       | 20%               | 25%                 | 50%             | 35%                          | 24%                  |
| Disease / condition status  | 27%   | 27%  | 30%       | 40%               | 50%                 | 0%              | 15%                          | 24%                  |
| Mortality                   | 8%    | 13%  | 10%       | 8%                | 13%                 | 0%              | 10%                          | 7%                   |
| Risk scores                 | 4%    | 10%  | 5%        | 4%                | 0%                  | 0%              | 10%                          | 4%                   |
| Behaviors                   | 3%    | 7%   | 20%       | 0%                | 0%                  | 0%              | 0%                           | 9%                   |
| Multiple                    | 15%   | 4%   | 20%       | 24%               | 0%                  | 50%             | 20%                          | 16%                  |
| Other                       | 9%    | 2%   | 5%        | 4%                | 13%                 | 0%              | 10%                          | 16%                  |
| <b>Study population</b>     |       |  |           |                   |                     |                 |                              |                      |
| Condition                   | 54%   | 57%  | 65%       | 40%               | 25%                 | 0%              | 75%                          | 67%                  |
| Demographic                 | 18%   | 15%  | 10%       | 24%               | 38%                 | 50%             | 0%                           | 13%                  |
| Organizational              | 15%   | 15%  | 25%       | 12%               | 25%                 | 0%              | 15%                          | 9%                   |
| Geographic                  | 13%   | 13%  | 0%        | 24%               | 13%                 | 50%             | 15%                          | 11%                  |
| <b>Children included</b>    | 21%   | 21%  | 0%        | 36%               | 50%                 | 0%              | 15%                          | 13%                  |
| <b>Level of measurement</b> |       |  |           |                   |                     |                 |                              |                      |
| Aggregate                   | 50%   | 50%  | 0%        | 80%               | 63%                 | 100%            | 45%                          | 16%                  |
| Individual                  | 29%   | 30%  | 100%      | 20%               | 38%                 | 0%              | 50%                          | 84%                  |
| Both                        | 21%   | 20%  | 0%        | 0%                | 0%                  | 0%              | 5%                           | 0%                   |

# Explore IADC Data Resources

 **IADC Portal** [Home](#) [I want to know more](#) Hello danhood

All Datasets

All

Secured Datasets

Public Datasets

Domains

Granularity

Stewardship

**American Community Census 5 year Estimates (US Census Bureau)**  
Census Bureau's "American Community Survey" is 5-Year periodic Estimate that include a wide range of Population, Economic, Housing and various Demographics / Social data sets. The 2017 Survey includes 643 Indiana specific data tables for items like Race, Poverty status, Language, Veteran status, etc. which can be grouped by Census Regions like: State, County, Census Tract, etc.

**Eviction Data (Eviction Lab)**  
Eviction Lab data contains Eviction counts and Rates plus Eviction filings combined with key Census data like Regions, Race, Income. Some data is estimated.


**Food Access Research Atlas (USDA)**  
Provided by USDA, This data set contains food access indicators like Distance, Vehicle access grouped by Census Tracts within a Region. This data set includes key Census attributes like Race, Income and Housing


**Indiana Hospital Discharges (Indiana Hospital Association)**  
Hospital discharge data is released each year by the Indiana State Department of Health. Inpatient data is aggregated by hospital, payer, APR-DRG (All Patients Refined Diagnosis Related Groups), MS-DRG (Medicare Severity Diagnosis Related Groups), principal diagnosis, and principal procedure is provided by year.

**Washington Post DEA Opioid Pill Database**  
The Washington Post sifted through nearly 380 million transactions from 2006 through 2012 that are detailed in the DEA's database and analyzed shipments of oxycodone and hydrocodone pills, which account for three-quarters of the total opioid pill shipments to pharmacies. The Post is making this data available at the county and state levels in order to help the public understand the impact of years of prescription pill shipments on their communities.

**Dataset Highlights**

- Granularity**  
Geographic level data: Census tract, zip code, county.
- Linking Attributes**  
GEOGRAPHIC can be Linked to GEOGRAPHIC data (linkages can only be made to geographic boundaries of the same size or larger)
- Coverage Years**  
Metadata is derived from 2017 ACS survey, however ACS data can be pulled from 2005-2017
- References**  
<https://www.census.gov/programs-surveys/acs>
- Updated: Yearly**

 **IADC Portal** [Home](#) [Back](#) [I want to know more](#) Hello danhood

**Indiana Network for Patient Care**  
Indiana Network for Patient Care Data set is a longitudinal view nearly all health care that is provided in Indiana and includes a full set of Clinical and Encounter information. Clinical data includes: Diagnoses, Results, Laboratory reports, Medications, Orders, Procedures, Textual Reports all associated with a given encounter. Contains over 18 million patients, 7 billion clinical observations, 1.1 billion encounter records, 290 million text reports, Patients from multiple institutions are matched / linked

**Dataset Highlights**

- Granularity**  
Person level data
- Linking Attributes**  
PERSON can be Linked to other PERSON data; PERSON can be Linked to GEOGRAPHIC data
- Coverage Years**  
Rich data from 2005 forward for most health organizations in Indiana and data back to 1980's and 90's for select health organizations
- References**  
<https://www.regenstrief.org/implementation/data-core/data-guide/>
- Updated: Yearly**
- Attributes**
  - DOMAIN\_IDS
  - ...

**Dataset Structure** **Domain**

**Dataset Structure** [View All](#)

Previous

Page 1 of 1

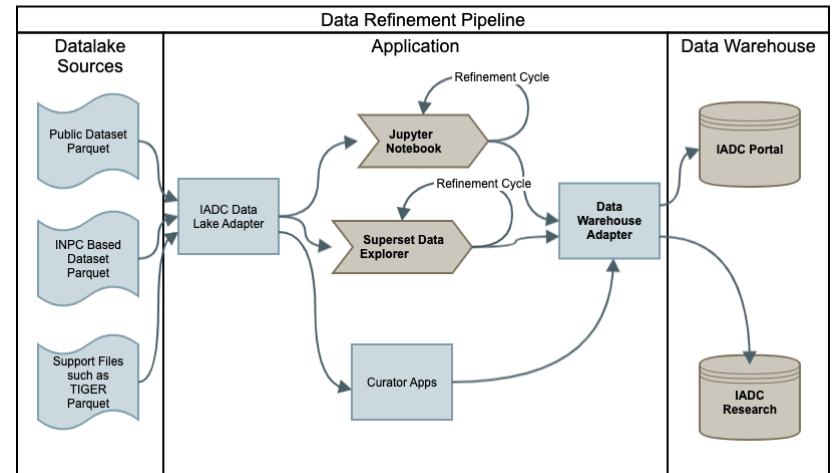
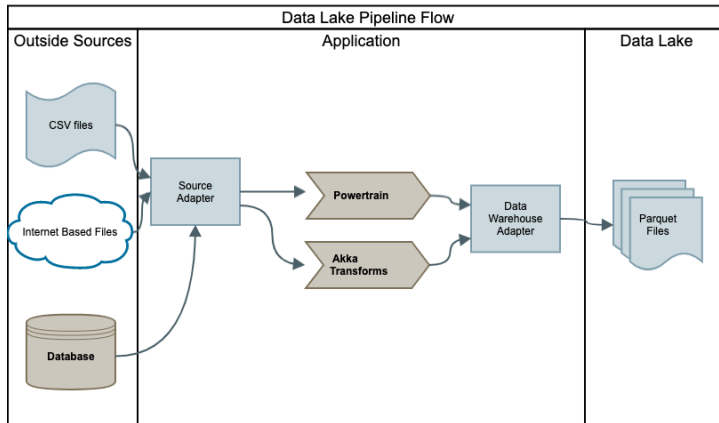
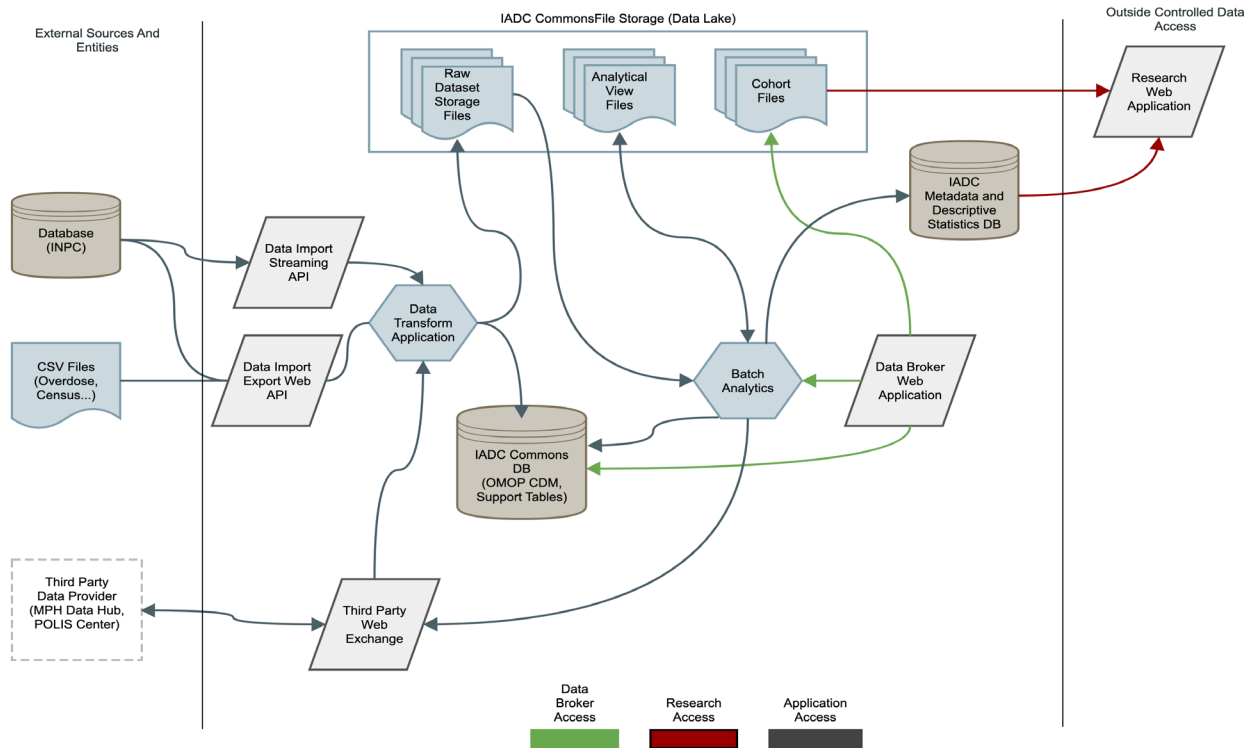
50 rows

Next

| Name                  | Domain Ids             | Vocabulary Ids   |
|-----------------------|------------------------|--|
| ▶ APPOINTMENT         | Appointments           | This table contains data on outpatient appointments that have ...      |
| ▶ ENCOUNTER           | Encounter              | This Data defines every "encounter" between a patient and a h...       |
| ▶ CLINICAL_VARIABLE   | Clinical Observations  | Clinical observations include labs, vitals, pathology, microbiolo...   |
| ▶ MEDICAL_ORDER       | Medical Orders         | Orders for any medical services: prescriptions, consultations, la...   |
| ▶ MULTIMEDIA_CONTENT  | Text Reports           | Contains a Textual / Narrative reports, visits/consults, radiology...  |
| ▶ PHARMACY_ORDER      | Medication Orders      | Detailed information about medication orders and fill events, D...     |
| ▶ PATIENT             | Patient Information    | This table contains information that is restricted to patients (as ... |
| ▶ PERSON              | People                 | This table contains Demographic information attributed to any ...      |
| ▶ PERSON_RELATIONSHIP | Relationships          | To show biologic relationships between 2 persons, such as mot...       |
| ▶ PARTICIPATION       | Provider Participation | This table will record how all clinical providers participated in a... |
| ▶ LOCAL_PROVIDER      | Clinical Provider      | Describes all providers of clinical care and includes demograph...     |
| ▶ LOCAL_USER          | System Users           | All users of the institution's EHR system.                             |
| ▶ DIAGNOSES           | Diagnoses              | A variety of patient diagnoses are stored and can often be link...     |



# Data Lake Development



# IADC Approaches to Governance

- Formation of a Governance Council early on
  - Included representatives of data partners and stakeholders
- Development of IADC Governance Charter
  - Purpose and vision, guiding principles, goals and objectives
- ***Key Objective: development of processes and agreements to enable data use***
- Founding Committee members:
  - Indiana MPH, IBRC, IHA, MCPHD, IU, IHIE, HC1, Polis Center, Regenstrief Institute

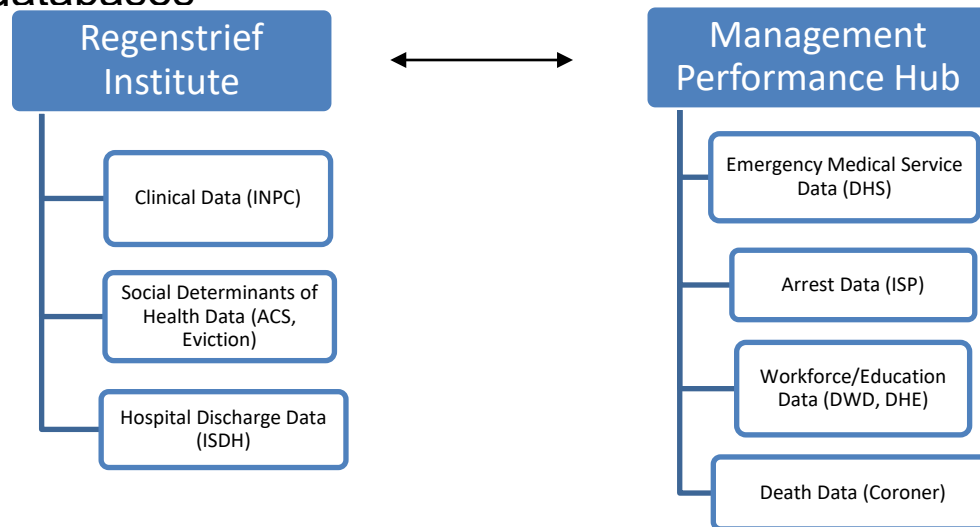
# Challenges

*Many...*

| Data                                       | Access                               | Processes                             | Sustainability            |
|--|--------------------------------------|---------------------------------------|---------------------------|
| Non-standardized data collection           | Varying levels of governance         | Aligning multiple stakeholders        | Motivation to participate |
| Ltd. SDoH collection at point-of-care      | Disparate data sources remain siloed | Lack of interconnectivity among orgs. | Competing efforts         |
| Data quality, completeness, representation | Use-case driven                      | Avoiding one-off's                    | Resources                 |
| Limitation of area level measurements      | Trust                                |                                       |                           |

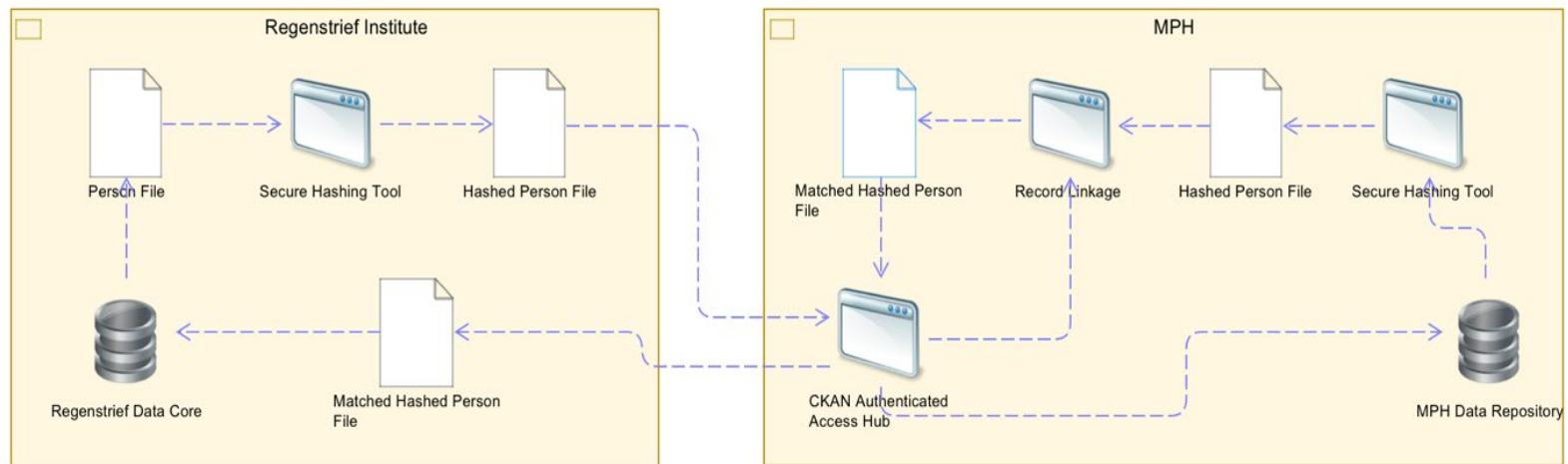
# Solutions (data access/governance/processes)

- Linking EMR and State level data
  - Enable a reproducible linkage process between persons derived from the INPC and persons existing in the MPH database
  - Enable a process for data exchange for persons existing in both databases
- RI and MPH executing an overarching DSA which streamlines data sharing between two organizations
  - Enables data flows for person level linkage
  - Enables access/storage of RI dataset in MPH enhanced research environment



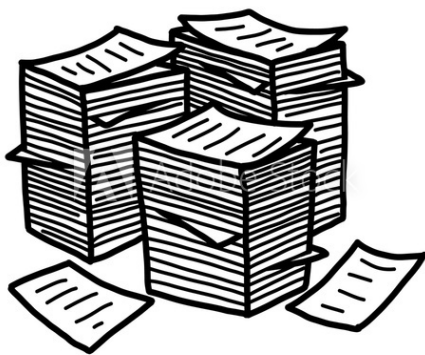
# Linking EMR and State Level Data

- Enabled privacy preserving record linkage to protect unique IDs
  - *First name, last name, DOB, Gender, Phone, Zip, SSN, Last 4 SSN*
- Both sides created a unique pseudonymous ID for each individual
  - *Maps to each organizations original data*
- Apply a locality sensitive hashing algorithm + secret SALT



# Solutions (data collection)

- Extracting SDoH concepts from free text clinical notes
  - Collection of SDoH in EMR is improving, but structured data is limited
  - Developing algorithms focused on
    - Housing Instability
    - Unemployment
    - Financial insecurity
    - History of incarceration
    - Transportation issues
  - Translation of unstructured indicators into structured SDoH flags



#177710578

Natural  
Language  
Processing



| Patient ID | Housing Instability | Unemployment | Trans. Issues | ... | Date      |
|------------|---------------------|--------------|---------------|-----|-----------|
| 100001     | -                   | -            | X             | -   | 1/23/2020 |
| 100002     | -                   | X            | X             | -   | 2/25/2020 |
| 100003     | X                   | X            | -             | -   | 4/1/2020  |

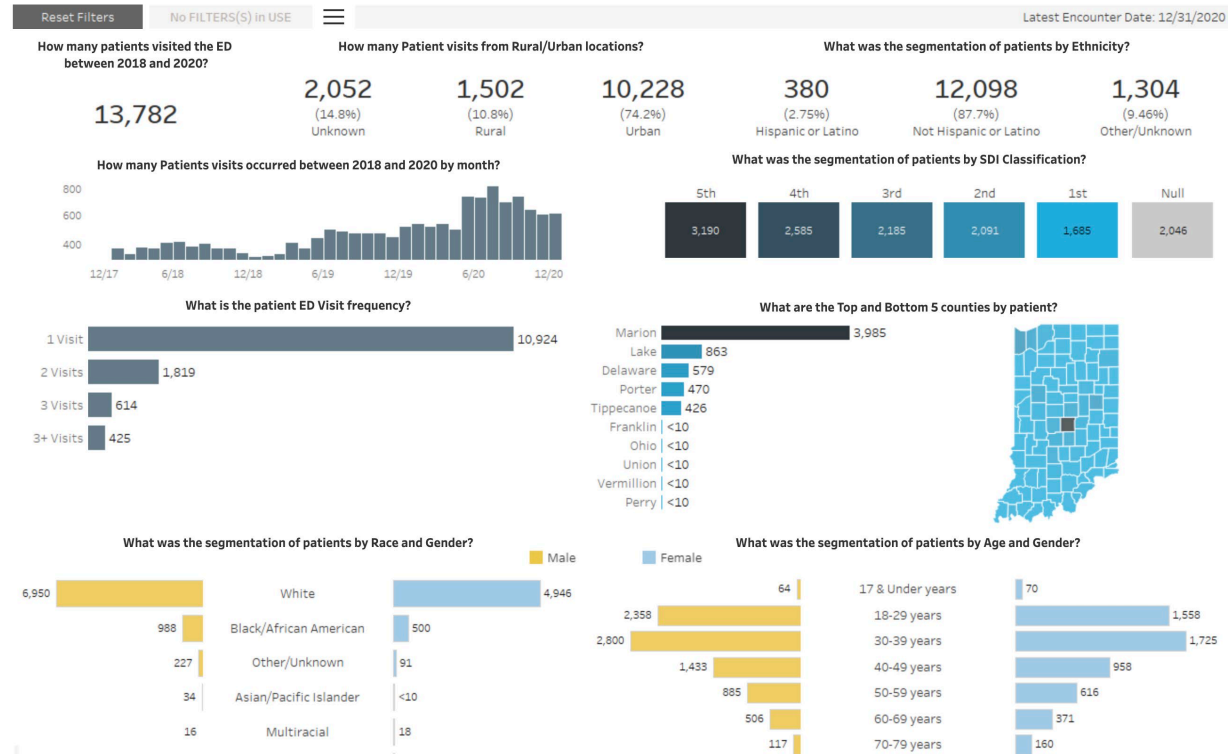


# Solutions (data curation)

- IADC is focused on developing curated cohorts for populations of interest
- Cohorts will form building blocks for integration with additional SDoH data
- IADC will be releasing a series of interactive dashboards

## Individuals with an ED visit related to an opioid use disorder

<< Go Back



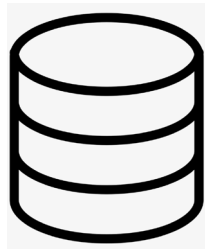
# IADC Supported Projects

- IADC Integrated Data Set Development RFA
  - Launched in Fall of 2019
  - 3 project teams chosen to support data set development
- Marion County Public Health Department-Overdose to Action Initiative (OD2A)
- Other Addictions Grand Challenge projects
- Mobile Crisis Assistance Team (MCAT)
- Natural Language Processing to identify SDoH concepts

# Next Steps (data access)

- Enable more streamlined access to priority data sources
  - Food access, criminal justice, medication assisted therapy
  - Define clear and reproducible processes for obtaining these types of data
  - Generate support from organizations who govern these data

- Enable more self service capabilities within the IADC infrastructure
  - Current processes are mediated through IADC team
  - Future goal would be to enable feasibility exploration of data sets



*Move from dataset creation to exploration to feasibility analysis*

# Next Steps (collaboration)

- Align with external groups working in the SDoH space
- For example: Gravity Project
  - Funded by RWJ, led by SIREN UCSF
  - Developing data standards for SDoH documentation in health care setting



# Next Steps (expand use cases)

- IADC is primarily focused on addictions related work
- Need for SDoH data is not limited to addiction
- Expand scope of data use to other public health crises

