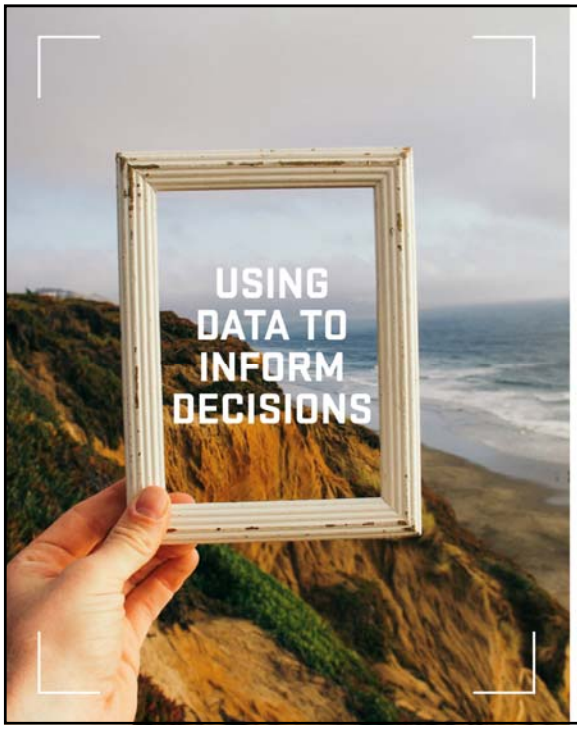
2

TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK

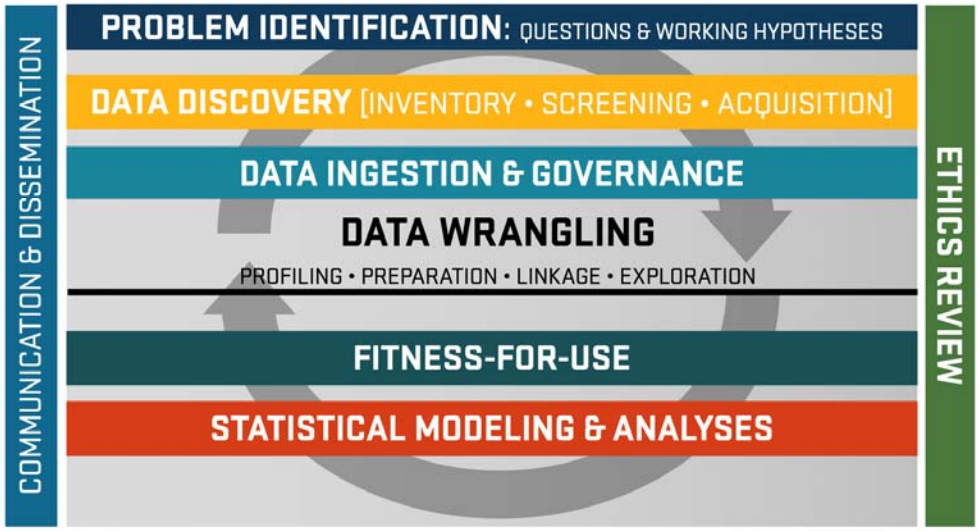
DATA DRIVEN DISCOVERY PROCESS

Adopted from materials provided by Dr. Sallie Keller and team at
The Social Decision Analytics Division,
Biocomplexity Institute, University of Virginia



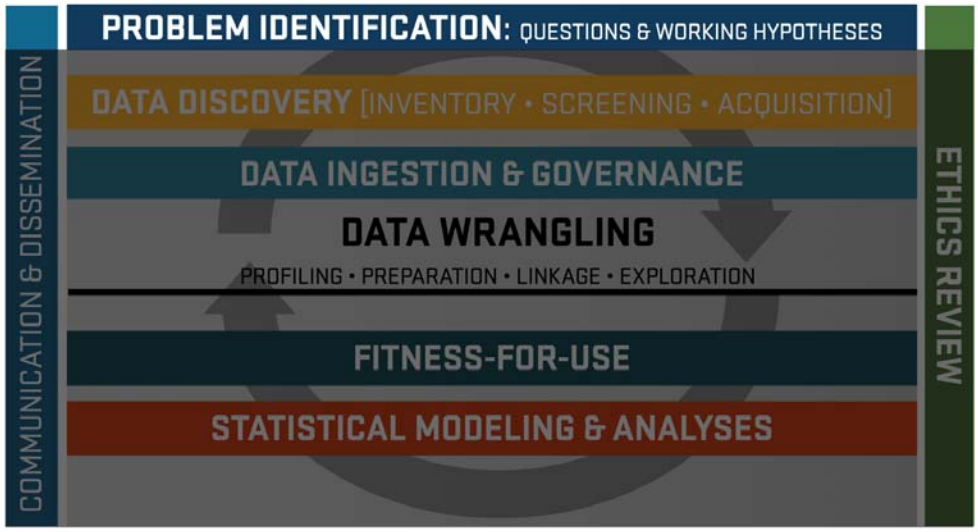
1. WHAT IS THE **PROBLEM?** WHAT ARE THE **QUESTIONS?**
2. WHAT **TYPE(S) OF DECISION** NEEDS TO BE MADE?
3. WHAT **DATA ARE AVAILABLE?** WHY SHOULD I **TRUST THE DATA?**
4. WHAT **TYPE OF ANALYSES** ARE NEEDED TO INFORM THE DECISION?
5. WHAT ARE THE **ETHICAL CONSIDERATIONS?**

KNOW WHERE YOU ARE IN THE DATA SCIENCE PROCESS



USING DATA TO INFORM DECISIONS

What is the **problem**? What are the **questions**?



5 TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK

EXAMPLES: USING DATA TO INFORM DECISIONS

ARLINGTON COUNTY, VIRGINIA

- **PROBLEM:** Improve quality of life and services while accelerating county's efficiency and resiliency
- **GOAL:** Develop data-informed situational awareness with the goal
- **1ST QUESTION:** Fire Chief: "Can we link Fire/EMS data end-to-end to recreate incidents to understand resource utilization?"

FAIRFAX COUNTY, VIRGINIA

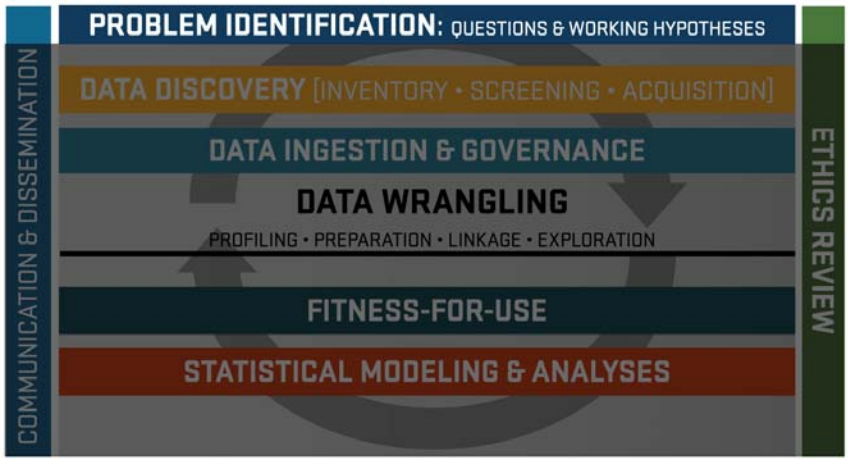
- **PROBLEM** - Identify barriers to economic mobility, e.g., income, jobs, health, education ...
- **GOAL:** Use local data to inform county-wide, long-term issues and establish baselines for future policy and program evaluation
- **1ST QUESTION** – Health and Human Services: "What are the neighborhood trends in youth obesity, access to healthy food, and physical activities?"

5 TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK

7

USING DATA TO INFORM DECISIONS

What is the **problem**? What are the **questions**?
What **type(s) of decision** needs to be made?

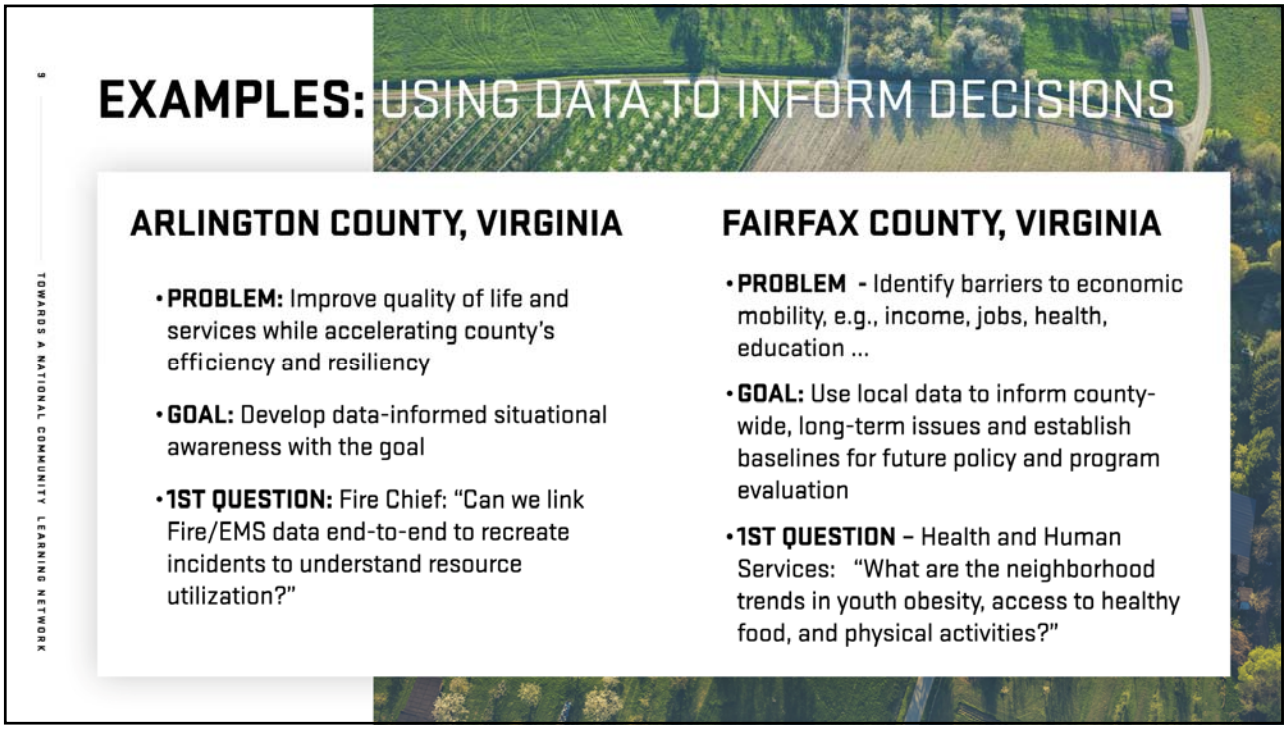


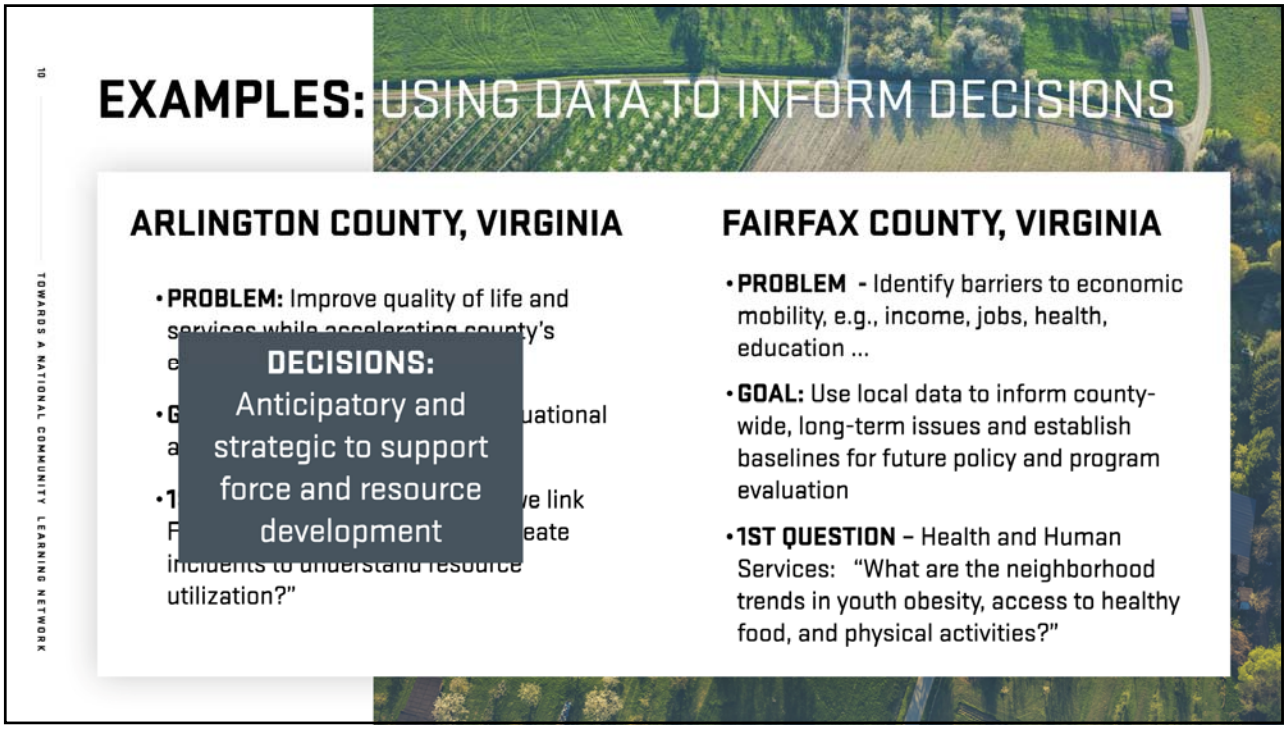
8

TYPES OF DECISIONS

- **STRATEGIC DECISION-MAKING** – inform and enrich understanding of problems and issues to plan for future operations
- **ANTICIPATORY DECISION-MAKING** – detect, identify, and warn of emerging issues and discontinuities
- **CURRENT (TACTICAL) DECISION-MAKING** – support ongoing actions and operations

Adapted from https://www.dni.gov/index.php/what-we-do/what-is-intelligence


9
TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK
EXAMPLES: USING DATA TO INFORM DECISIONS
ARLINGTON COUNTY, VIRGINIA
• PROBLEM: Improve quality of life and services while accelerating county's efficiency and resiliency
• GOAL: Develop data-informed situational awareness with the goal
• 1ST QUESTION: Fire Chief: "Can we link Fire/EMS data end-to-end to recreate incidents to understand resource utilization?"
FAIRFAX COUNTY, VIRGINIA
• PROBLEM - Identify barriers to economic mobility, e.g., income, jobs, health, education ...
• GOAL: Use local data to inform county-wide, long-term issues and establish baselines for future policy and program evaluation
• 1ST QUESTION – Health and Human Services: "What are the neighborhood trends in youth obesity, access to healthy food, and physical activities?"


10
TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK
EXAMPLES: USING DATA TO INFORM DECISIONS
ARLINGTON COUNTY, VIRGINIA
• PROBLEM: Improve quality of life and
DECISIONS:
Anticipatory and strategic to support force and resource development
utilization?"
FAIRFAX COUNTY, VIRGINIA
• PROBLEM - Identify barriers to economic mobility, e.g., income, jobs, health, education ...
• GOAL: Use local data to inform county-wide, long-term issues and establish baselines for future policy and program evaluation
• 1ST QUESTION – Health and Human Services: "What are the neighborhood trends in youth obesity, access to healthy food, and physical activities?"

EXAMPLES: USING DATA TO INFORM DECISIONS

ARLINGTON COUNTY, VIRGINIA

- **PROBLEM:** Improve quality of life and services while accelerating county's e
- G uational a
- 1 ve link F eate incidents to understand resource utilization?"

DECISIONS: Anticipatory and strategic to support force and resource development

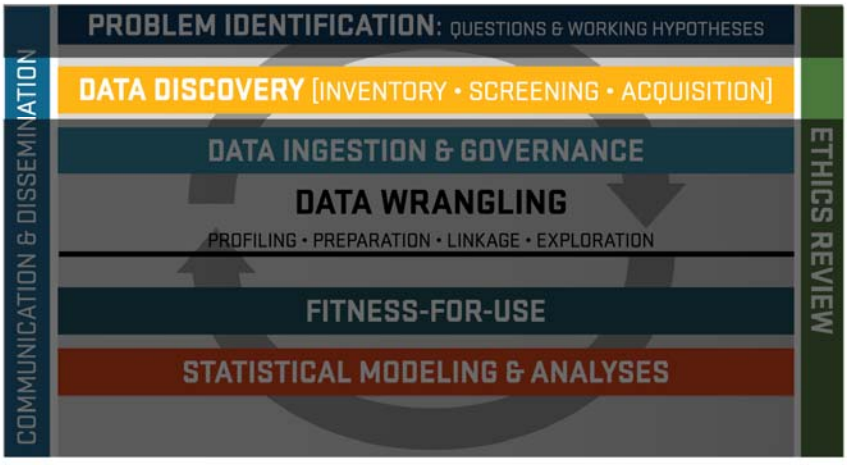
FAIRFAX COUNTY, VIRGINIA

- **PROBLEM** - Identify barriers to economic mobility, e.g., income, jobs, health, educ
- GOA y-wide base am evalu
- 1ST Services: "What are the neighborhood trends in youth obesity, access to healthy food, and physical activities?"

DECISIONS: Anticipatory and strategic to guide policy development and planning

USING DATA TO INFORM DECISIONS

What is the **problem**? What are the **questions**?
What **type(s) of decision** needs to be made?
What **data are available**? Why should I **trust the data**?

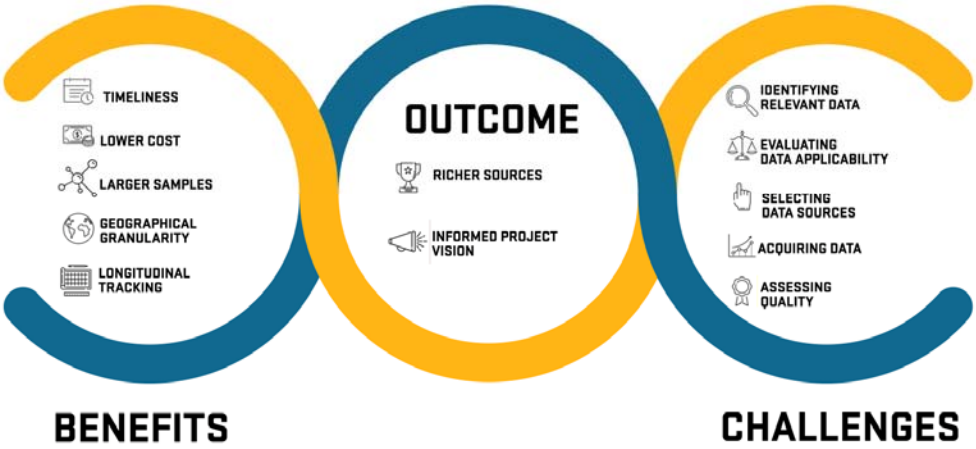


DATA DISCOVERY

PROCESS & DEFINITIONS

- **DATA DISCOVERY** is the open-ended and continuous process by which potential datasets are identified
- **DATA INVENTORY** refers to the broadest, most far-reaching 'wish list' of information pertaining to your 'big' question
- **DATA SCREENING** is an evaluative process by which eligible datasets are sifted from the larger pool of candidate datasets
- **DATA ACQUISITION** is the method of physically acquiring the data, which can involve cooperative or purchase agreements, web-scraping, and file-transfer protocols

DATA DISCOVERY: BENEFITS & CHALLENGES



Accessing Alternative Sources of Data for Education Research. Ziemer, Shipp, et al, 2016.

DATA INVENTORY

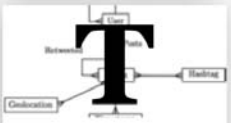
CONSIDER ALL RELEVANT DATA SOURCES:
LOCAL, STATE, FEDERAL

DESIGNED DATA

- CENSUS SURVEYS
- CDC SURVEYS
- BRFSS SURVEYS

ADMINISTRATIVE DATA

- BIRTH RECORDS
- ER VISIT RECORDS
- WASTE MANAGEMENT



OPPORTUNITY DATA

- GPS
- SOCIAL MEDIA

PROCEDURAL DATA

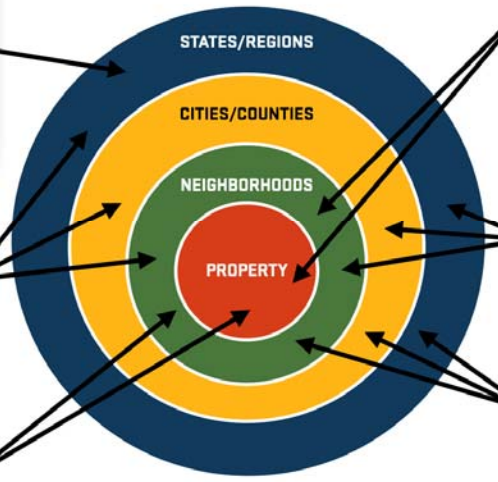
- BILLS
- SUPREME COURT RULINGS
- LEGAL BRIEFS

DATA INVENTORY: CONSIDER ALL RELEVANT DATA DOMAINS

BROADBAND FACTORS
- Availability
- Number of providers
- Type
- Speed
- Latency
- Consistency
- Adoption rate

NEIGHBORHOOD FACTORS
- Schools
- Safety/Crime Indicators
- Cultural Institutions
- Parks & Recreation
- Healthcare
- Public Transportation

COMMUNITY DEMOGRAPHICS
- Age Distribution
- Education Distribution
- Income/Wealth Indicators
- Racial/Ethnic Composition



PROPERTY FACTORS
- Value (land, improvement)
- Location
- Age (actual, effective)
- Ownership/Rental
- Size (acreage/square footage)
- Public Infrastructure (roads/utilities)

HOUSING FACTORS
- Construction Activity
- Zoning/Land Use Restrictions
- # of Homes For Sale
- # of Buyers

DEVELOPMENT FACTORS
- Unemployment
- Business & Entrepreneurship
- GDP (AG, MFG., FORESTRY, ...)
- Income Growth
- Inflation Rates
- Commercial Amenities
- Infrastructure Plans

DATA DISCOVERY

LET'S TRY IT FOR OUR PROBLEM!

ISSUE

- How can the Iowa Department of Public Health better monitor drug use and its impacts on individuals and families throughout our state?

STEP ONE: DATA INVENTORY

- Each team will brainstorm dataset 'wish lists' related to surveilling substance use among individuals and families in Iowa

STEP TWO: DATA SCREENING

- Each team will brainstorm screening criteria that will be used to assess data
- Teams will evaluate the data identified in the Data Inventory to determine which datasets to keep or shelve

TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK

BRAINSTORM ACTIVITY

45 MINUTES—RECORD ALL IDEAS!

What **information** and **data** would be useful to help IDPH monitor drug use and its impacts on individuals and families in Iowa?

DATA ENVIRONMENT: SUBSTANCE USE AMONG IOWA FAMILIES

What themes emerge from problem elaboration?

PROBLEM

How can the Iowa Department of Public Health better monitor drug use and its impacts on individuals and families throughout the state?

THEMES

- Social context of drug use includes, family, community, church, and occupation
- Triggers for drug use include social isolation and stigma; economic vulnerability; and trauma and coping
- Mental and physical health tied to drug use

Who is engaged in or impacted by this work?

ENGAGED

- Agency partners
- Community leaders
- Treatment centers
- Needle exchange network
- Program clients
- Medical professionals
- Criminal justice system
- You!

IMPACTED

- Agency partners
- Community members
- Treatment centers
- Needle exchange network
- Program clients
- Medical professionals
- Criminal justice system
- Employers
- Children, families, adults

What internal or external factors affect this space?

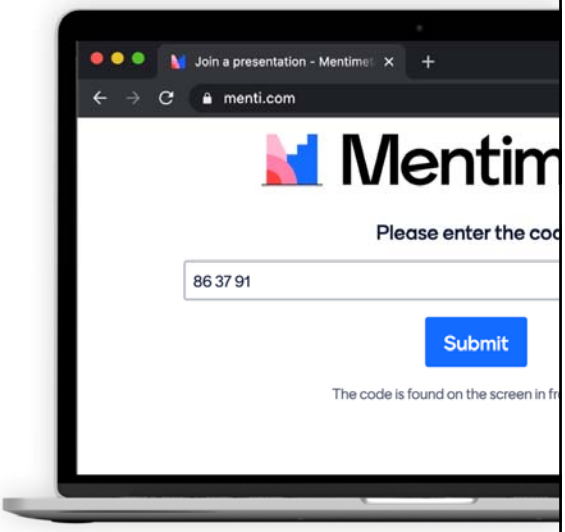
- Laws and regulations
- Programs and formal supports
- Informal supports
- State or community resources
- Stakeholder preferences
- Individual and family characteristics

What is the desired outcome of this process?

Are any patterns emerging? Do these suggest the type of data domains you should be considering?

HOW WOULD YOU USE A SUBSTANCE ABUSE MONITORING SYTEM TO SUPPORT PROGRAMMING OR POLICY EFFORTS?

GO TO www.menti.com
ENTER CODE 86 37 91

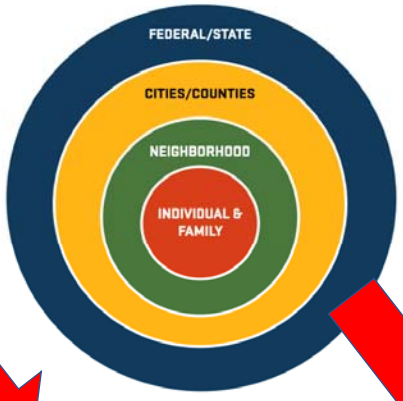


DATA DOMAINS: SUBSTANCE USE AMONG IOWA FAMILIES

DOMAIN: SUBSTANCE ABUSE & HEALTH
- Use of opioids, methamphetamines, marijuana, alcohol, cigarettes/e-cigarettes
- Drug related mortality
- Drug related crime
- Drug distribution
- Mental health
- Physical health
- Risk taking
- Deaths of despair

DOMAIN: SOCIAL ISOLATION
- Community characteristics
- Church & religious involvement
- Individual & community social capital
- Community & political engagement

DOMAIN: TRAUMA & COPING
- Adverse Childhood Experiences (ACES)
- Incarceration, probation, juvenile detention
- Mortality
- Crime



DOMAIN: FAMILY & INDIVIDUAL CHARACTERISTICS
- Family characteristics
- Child/teenager characteristics
- Household characteristics
- Births and fertility
- Sexual Orientation and gender
- Immigration status
- Language

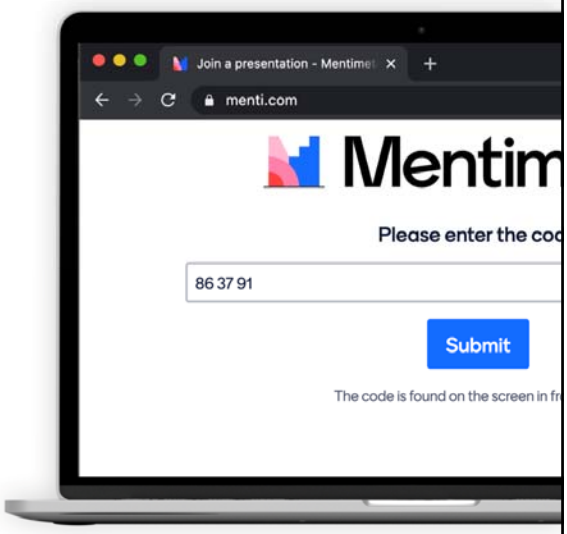
DOMAIN: ECONOMIC VULNERABILITY
- Occupation
- Income
- Housing
- Employment
- Community resources
- Economic system
- General economic wellbeing

DOMAIN:

DOMAIN:

WHAT OTHER INFORMATION DOMAINS SHOULD WE CONSIDER TO HELP US MONITOR 'HOT SPOTS' OF SUBSTANCE USE OR ITS EFFECTS?

GO TO www.menti.com
ENTER CODE 86 37 91

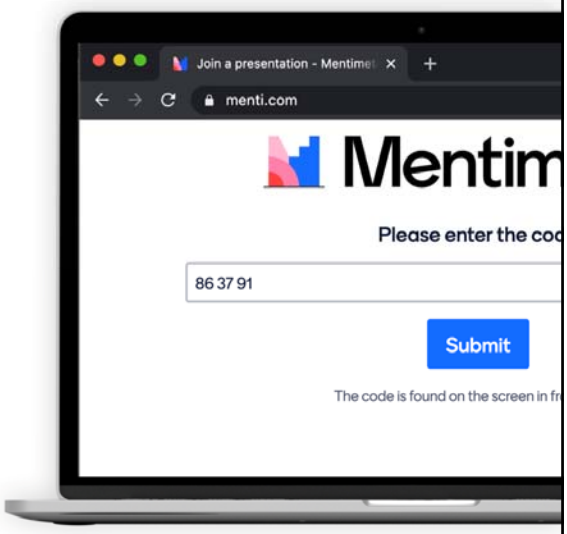


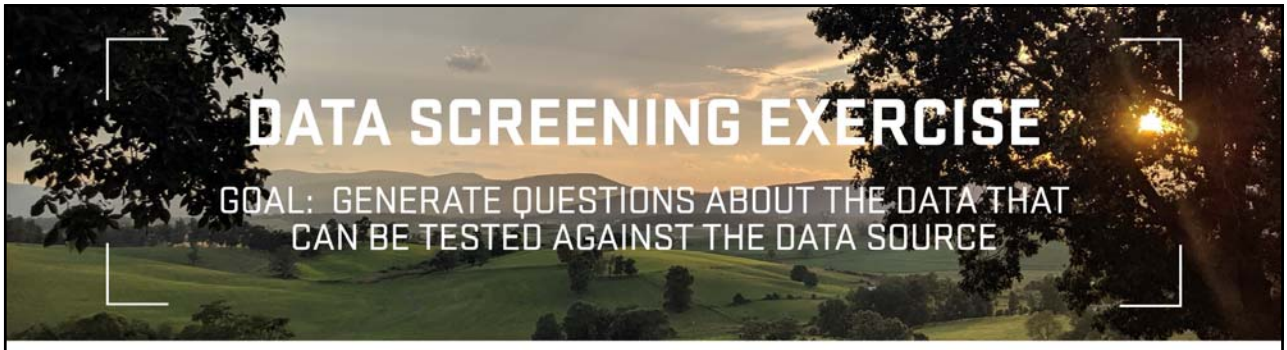
DATA INVENTORY: SUBSTANCE USE AN G IOWA FAMILIES

SUBSTANCE USE & HEALTH	FAMILY & INDIVIDUAL DEMOGRAPHICS	YOUR CHOICE
SOCIAL ISOLATION	ECONOMIC VULNERABILITY	YOUR CHOICE
		TRAUMA & COPING

WHAT DATA SETS DO YOU KNOW OF THAT WOULD HELP US MONITOR SUBSTANCE USE & ITS EFFECTS ON INDIVIDUALS & FAMILIES?

GO TO www.menti.com
ENTER CODE 86 37 91





BRAINSTORM ACTIVITY
15 MINUTES–RECORD ALL IDEAS!

What information do you need **ABOUT** the data to determine its usefulness?

IDENTIFY CRITERIA TO DETERMINE WHICH DATA TO "KEEP" OR "SHELVE"

DATA SCREENING: DEVELOP PROJECT SPECIFIC CRITERIA

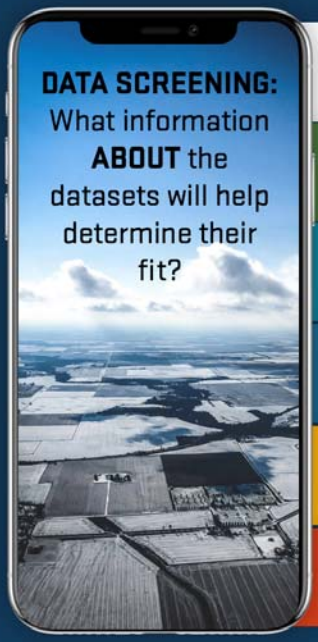
WHAT INFORMATION **ABOUT** THE DATASETS WILL HELP YOU DETERMINE THEIR FIT?

HOW FINE GRAINED A GEOGRAPHY IS NEEDED?

HOW FREQUENTLY ARE THE DATA COLLECTED?

WHAT IS THE DATA QUALITY? ARE THERE GAPS OR OTHER CONCERNS?

HOW DIFFICULT ARE THE DATA TO ACCESS? ARE LEGAL AGREEMENTS REQUIRED? IS THERE A FEE FOR USE?



1
- Does it cover at least every county in Iowa? Is there any greater geographic granularity? E.g., Yes/No and describe
- **SCREENING**: Determine what geography you need, such as county, neighborhoods, zip codes, and/or informal shapes.

2
- Is it current? How far back does it go? E.g., Yes/No & year range
- **SCREENING:** Does the data cover the time frame of interest?

3
- How are the data accessed? Are they open? Are there legal, regulatory, or administrative restrictions on accessing the data source? Cost? Is it one-time, annual, or a project-based payment? E.g., Yes/No
- **SCREENING:** What is shareable now? What is sharable in the future?

4
- Are there any serious gaps or concerns with this dataset? E.g., Yes/No
- **SCREENING:** Does the data have enough quality information to help answer the question at hand?

5
- Does the data cover recent events, climate, economy, political/legislative? Does it cover the population you need, such as non-English speakers? Is the sample size sufficient? E.g., Yes/No
- **SCREENING:** Does the data include the key features you need?

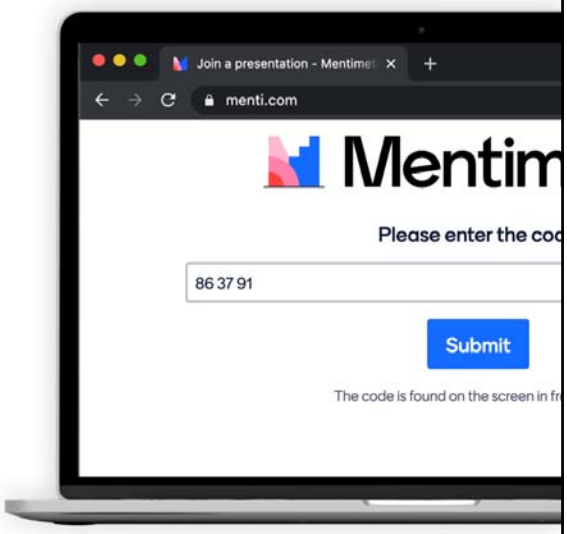
6
- Is there a cost associated to buy the dataset, labor, equipment, or other technical resources? How are additional costs factored in? Are there potential resources or funding sources for these? E.g., Yes/No
- **SCREENING:** Is it worth the investment?

28

TOWARDS A NATIONAL COMMUNITY LEARNING NETWORK

WHAT INFORMATION ABOU THE DATA SETS WILL HELP YOU DETERMINE THEIR FIT?

GO TO www.menti.com
ENTER CODE 86 37 91



DATA SCREENING: SUBSTANCE USE AMONG IOWA FAMILIES

DOES/IS THE DATA SET IN QUESTION:	DATASETS		
Cover the geography of interest?			
Cover time frame of interest?			
Shareable now? In the future?			
Of high enough quality to use?			
Include the necessary population, sample, or other			
Worth the investment?			

DESCRIBE ADDITIONAL ITEMS HERE

DATA SCREENING: SUBSTANCE USE AMONG IOWA FAMILIES

DOES/IS THE DATA SET IN QUESTION:	DATASETS		
Cover the geography of interest?			
Cover time frame of interest?			
Shareable now? In the fu			
Of high enough qu use?			
Include t ssary population ple, or other			
W investment?			

WRITE THE NAMES OF 3 DATA SETS HERE

FILL IN THE ANSWERS!

WE APPRECIATE FEEDBACK ON HOW TO ADDRESS THE IMPORTANT QUESTION OF BETTER MONITORING SUBSTANCE USE AND ITS EFFECTS IN IOWA

CONTACT:

CASS DORIUS
CDORIUS@IASTATE.EDU
SALLIE KELLER
SAK9TR@VIRGINIA.EDU