



# **New Tools for Real-Time Coding of Industry and Occupation in Field Surveys: An Approach Using CDC's Epi Info Software**

**Matthew Groenewold, PhD, MSPH**

**Epidemiologist, Division of Field Studies and Engineering**

Seventy-fourth Meeting of the NIOSH Board of Scientific Counselors

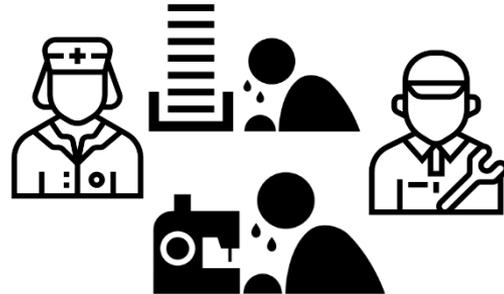
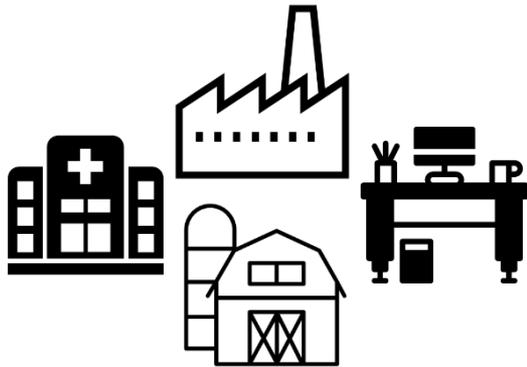
April 28, 2020

# Overview

- Why do we care about collecting data on industry and occupation in field surveys?
- Why is the ability to translate free text industry and occupation to standard codes in real time important?
- What are the challenges?
- How is NIOSH attempting to overcome these challenges (with a live demonstration)?
- What are the opportunities for BSC members and others to contribute?

# Industry and Occupation

- Basic demographic descriptors of working (or formerly working) people
- Should be included in all public health datasets
- Required for occupational health surveillance and research



# Collecting and Recording Industry and Occupation Data

- Usually collected as descriptions in free-text fields
  - **Industry:**
    - “What kind of business or industry do you work in? (for example, hospital, elementary school, clothing manufacturing, restaurant)”
  - **Occupation:**
    - “What kind of work do you do? (for example, registered nurse, janitor, cashier, auto mechanic)”
- Must be coded to be analyzed
  - Standard classification systems
    - North American Industry Classification System (NAICS)
    - Standard Occupational System (SOC)
    - US Census Bureau



# Collecting and Recording Industry and Occupation Data

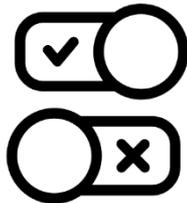
- Normally batch-coded after data collection
  - Impediment to use of industry and occupation data in field surveys
    - Outbreak investigations
    - Disaster response
  - Impediment to use of industry and occupation data in surveillance systems
    - Cases recorded one-by-one
    - Initial data entry by local, state health departments



# Collection of Industry and Occupation Data in the Field: Alternative Approaches

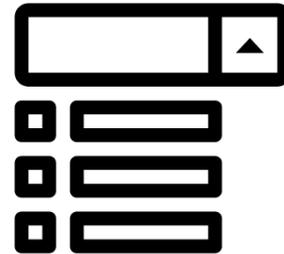
- **Directly specify critical/relevant occupations**

- “Are you a healthcare worker?” (Yes/No)
- “Are you a food handler?” (Yes/No)
- Limited utility
  - Presupposes single preexisting hypothesis



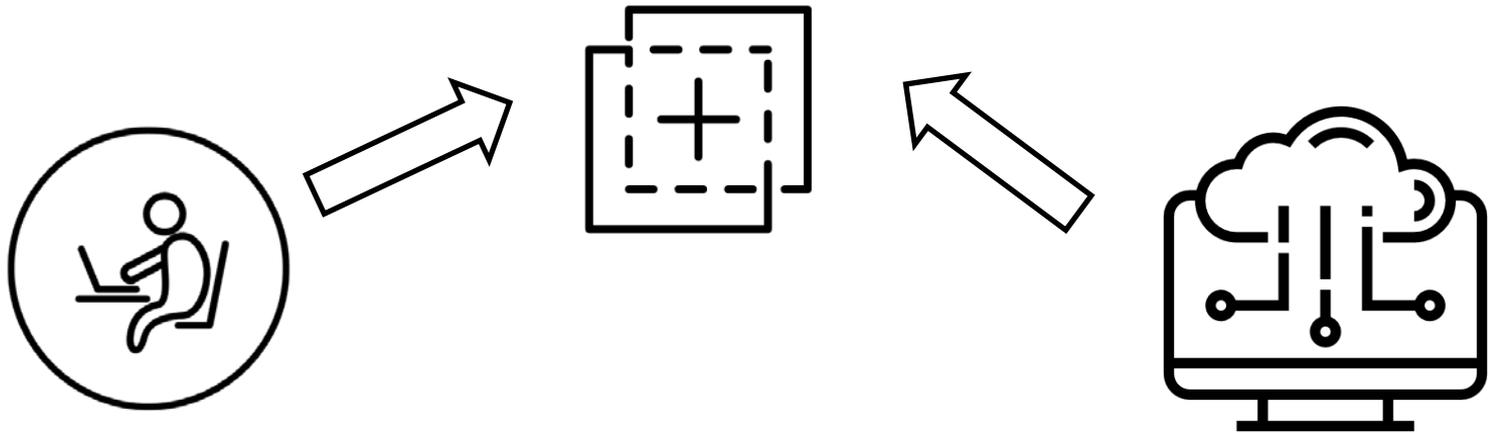
- **Restrict responses to pre-specified categories using dropdown menus**

- Categories may be too broad to be useful



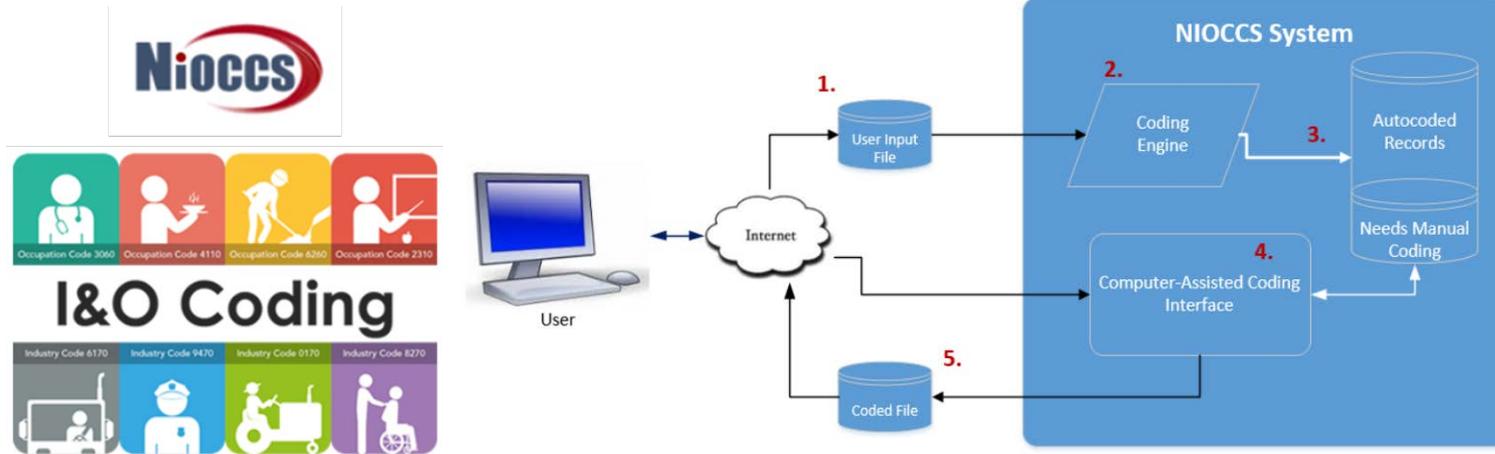
# Real-Time Coding of Industry and Occupation During Field Data Collection

- Ability to code industry and occupation data in real-time makes collection more feasible
- Integration of computerized coding capability with established field data collection tools enables real-time industry and occupation coding



# NIOCCS: The NIOSH Industry & Occupation Computerized Coding System

- NIOCCS provides on-demand, web-based single record coding capability



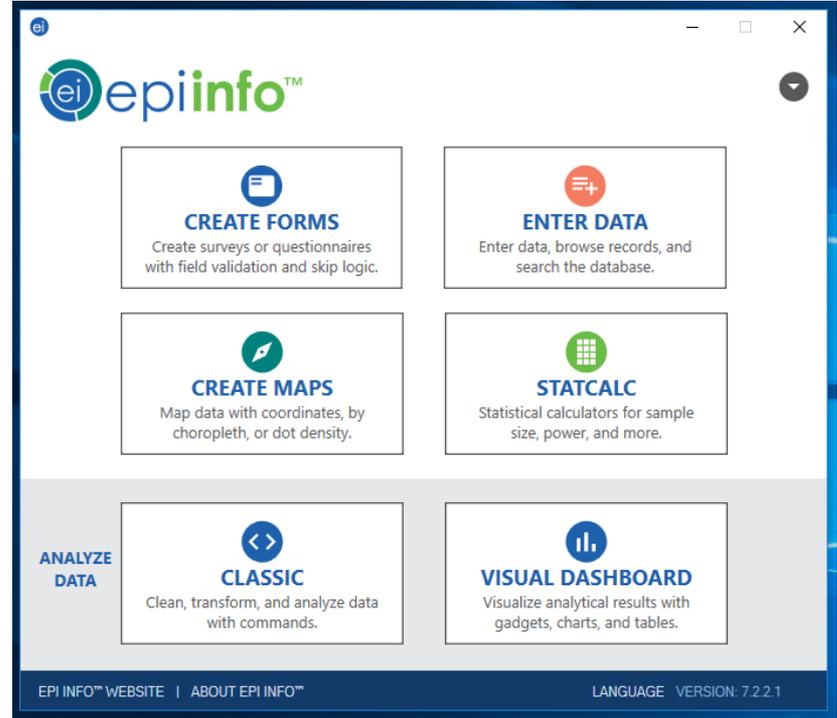
# Epi Info

- Suite of free data collection, management, analysis and visualization software tools designed specifically for the public health community
- Used throughout CDC, domestically and internationally
- Widely used by
  - Field epidemiology
  - Emergency preparedness and response
  - Reportable disease surveillance
  - State and local health departments



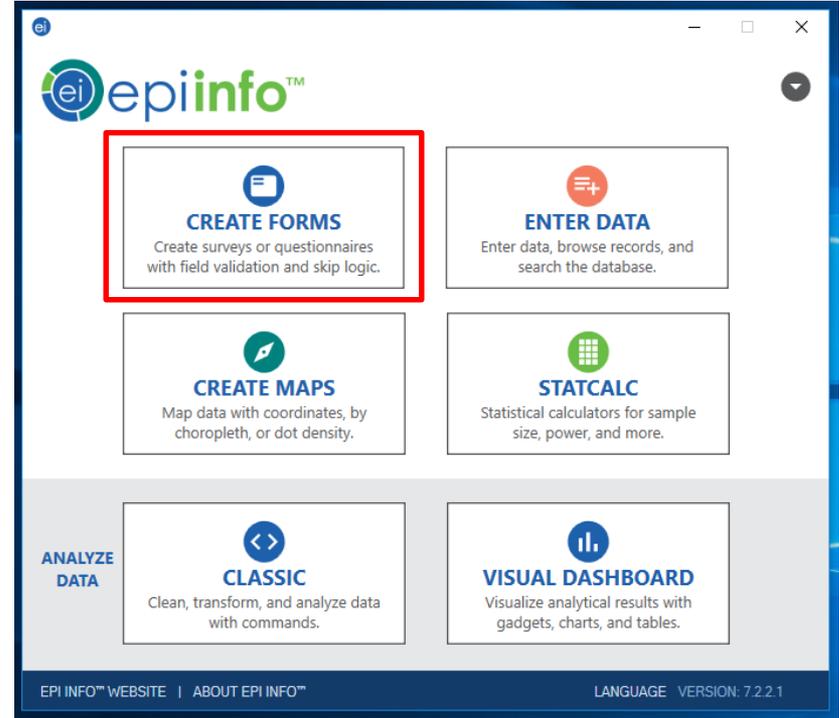
# Epi Info

- Form Designer module
- Enter module
- Check code
- Templates



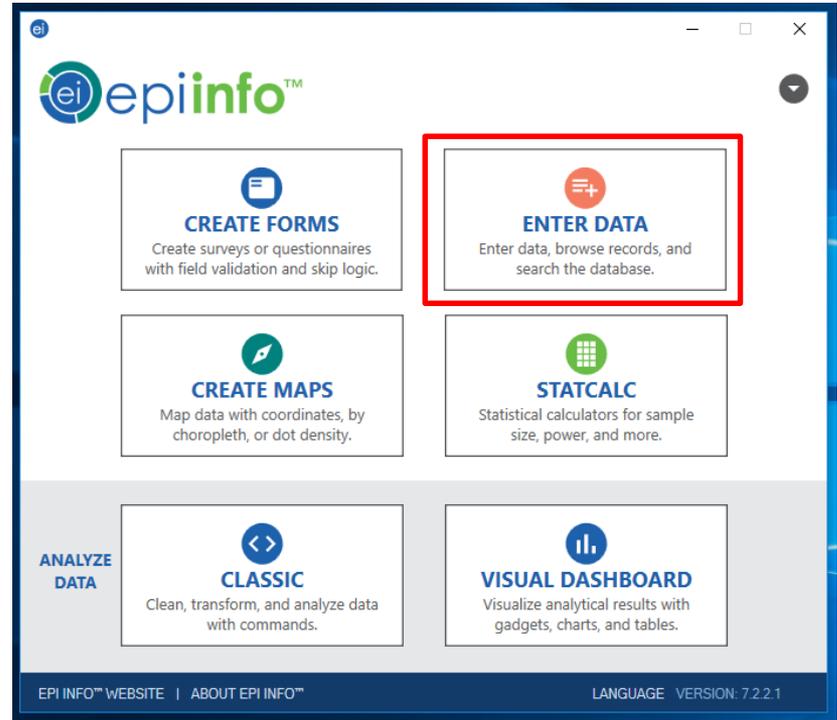
# Epi Info

- Form Designer module
- Enter module
- Check code
- Templates



# Epi Info

- Form Designer module
- Enter module
- Check code
- Templates



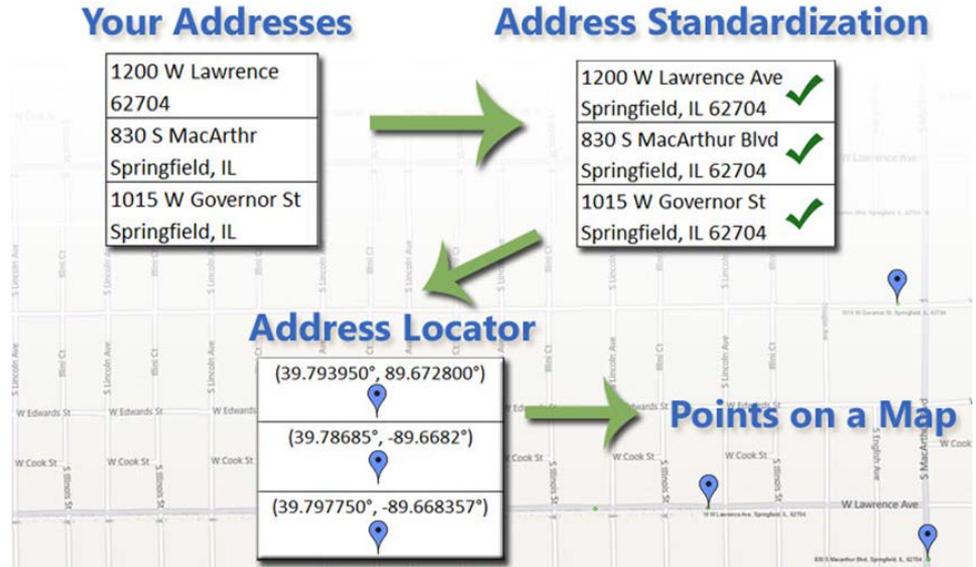
# Epi Info

- Form Designer module
- Enter module
- Check code
- Templates



# The Geocoding Analogy

- Geocoding process analogous to I & O coding
  - Input is free-text field
    - Cannot be analyzed
  - Output generated by algorithm
  - Results in analyzable, coded data
  - Typically batch-coded after data collection



# The Geocoding Analogy: Epi Info

- “Geocode” check code program allows Epi Info users to get map coordinates based on free-text address data during data collection

The diagram illustrates the geocoding process in Epi Info. It shows a 'Get Coordinates' button being clicked, which opens a 'Geocode Results' dialog box. The dialog box displays the geocoded address, confidence, quality, latitude, and longitude, along with an 'Accept' button and a source link.

**Initial State (Top Left):**

Address	Latitude
1853 Gresham Rd Louisville, KY 40205	
	Longitude
<input type="button" value="Get Coordinates"/>	

**Geocode Results Dialog (Middle Right):**

The geocoding service returned the following coordinates:

Address:	<b>1853 Gresham Rd, Louisville, KY 40205</b>
Confidence:	<b>High</b>
Quality:	<b>Good</b>
Latitude:	<b>38.21744</b>
Longitude:	<b>-85.69476</b>

Source: <http://www.microsoft.com/maps/>

**Final State (Bottom Left):**

Address	Latitude
1853 Gresham Rd Louisville, KY 40205	38.21744
	Longitude
	-85.69476
<input type="button" value="Get Coordinates"/>	

# The Geocoding Analogy: Epi Info

- “Geocode” program accessed in check code editor
- Pre-defined “Geo\_Location” template available in Form Designer module

The image displays two overlapping windows from the Epi Info software. The top window, titled "2. Add Command To Field Block", shows a list of commands including Assign, AutoSearch, Call, Clear, Define, Dialog, Disable, Enable, Execute, Geocode, GoTo, Help, Hide, Unhide, FieldSelector, Highlight, Unhighlight, and If. The bottom window, titled "Check Code Editor - [ Survey ]", shows a menu with File, Edit, Fonts, and Tools. The main text area contains the following code:

```
✓ Validate Check Code Close Save Print
/*
1. Choose a block from the upper right list to creat
2. Select a command to add to the block from the low
All Check Code commands must be within a block.
*/
/*
Check Code copied from Geo_Location.xml sample te
*/
Field GetCoordinates
  Click
    GEOCODE Address , Latitude , Longitude
  End-Click
End-Field
```

To the right of the Check Code Editor is a tree view of the Form Designer module. The tree structure is as follows:

- Codes
  - Relate
  - Group
- Templates
- Fields
  - Demographics
  - Diagnosis
  - Geo\_Location
  - Hospital Code Table Exam
  - Medical\_Facility
  - States
- Pages

# New Tools in Epi Info



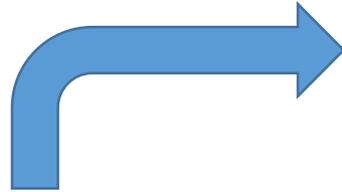
- “Code Industry and Occupation” check code program and template for Epi Info
  - Integrates industry and occupation coding capability with Epi Info data entry tools
  - Free-text to Census industry and occupation codes
  - Real-time
  - Record-by-record basis

# Function Overview (Original)



User enters subject's open-ended, free-text descriptions of industry and occupation in Epi Info's Enter module

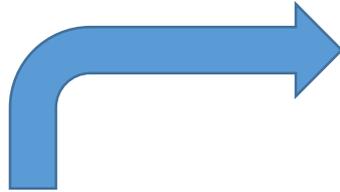
# Function Overview (Original)



User runs the “Code Industry and Occupation” check code program in Epi Info



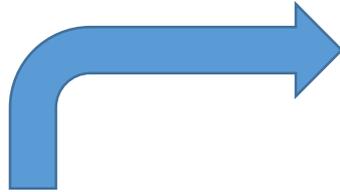
# Function Overview (Original)



"Code Industry and Occupation"  
program presents text data in  
industry and occupation fields to  
NIOCCS I & O Single Record  
Coding service via internet  
connection



# Function Overview (Original)



NIOCCS produces a coding solution



# Function Overview (Original)



Epi Info  
Check Code  
"Code  
Industry and  
Occupation"

NIOCCS output is displayed to user in an Epi Info dialog box

**Code Industry and Occupation**

Possible Matches:

Industry	Occupation
Confidence Level: 100% Code: 822 Code: 822 RESTAURANTS AND OTHER FOOD SERVICES	Confidence Level: 100% Code: 8132 Code: 8132 FOOD AND BEVERAGE STORES, RESTAURANT, SOURCE, AND CATERING

Accept

Accept



# Function Overview (Original)

Epi Info  
Check Code  
"Code  
Industry and  
Occupation"

User accepts or rejects NIOCCS code assignments, populating I & O code and title fields in the Epi Info project database



**Code Industry and Occupation**

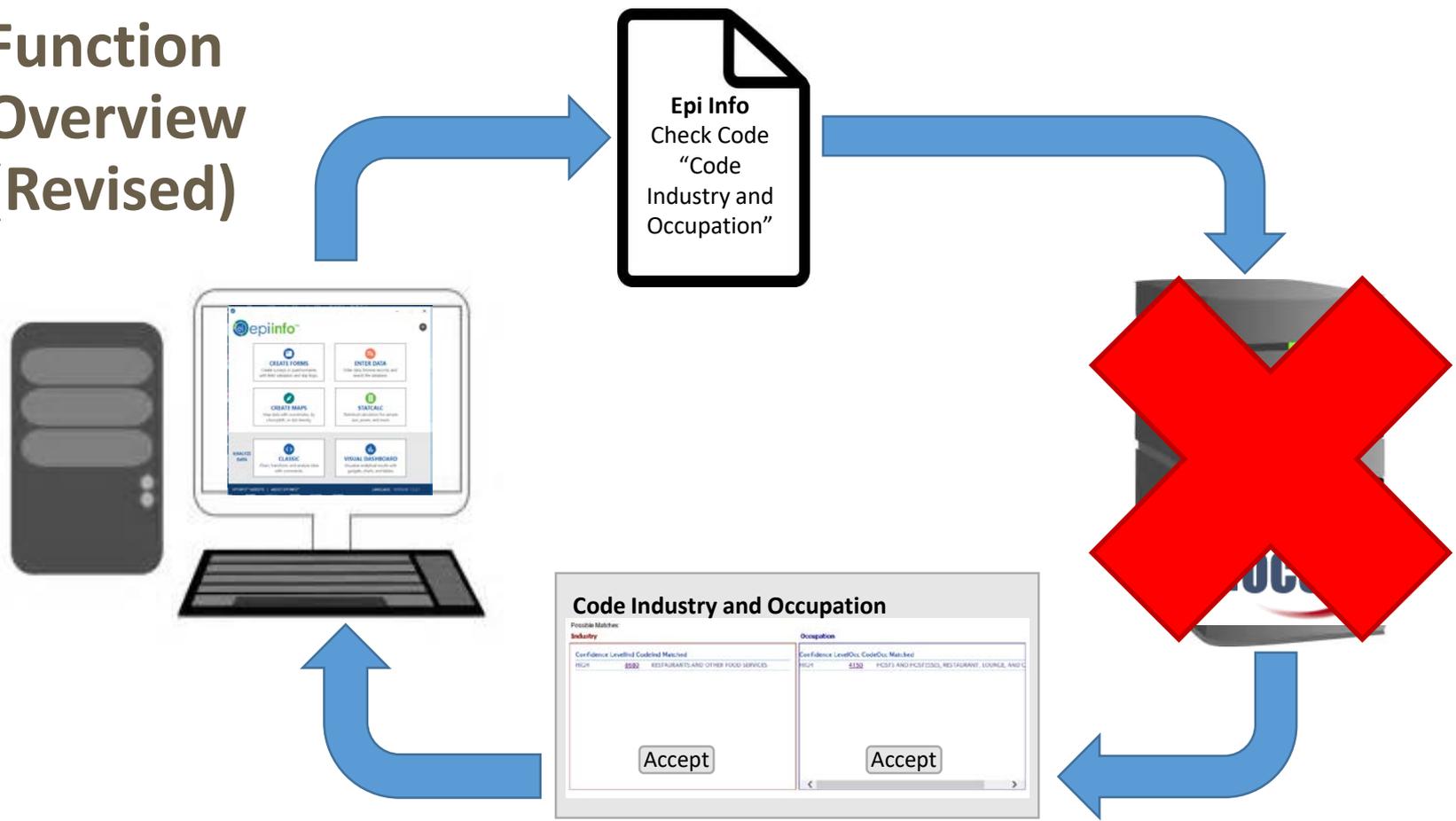
Probable Matches:

Industry	Occupation
Confidence Level: 0.80 Code: 8220 RESTAURANTS AND OTHER FOOD SERVICES	Confidence Level: 0.80 Code: 8120 FOOD AND BEVERAGE SERVICE, RESTAURANT, LOUNGE, AND C...

Accept

Accept

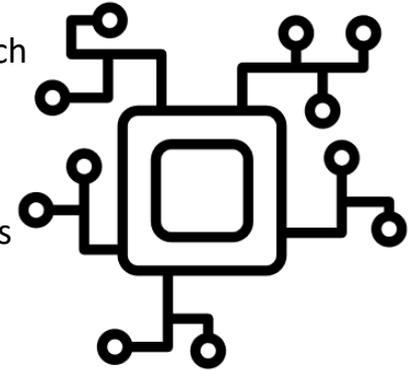
# Function Overview (Revised)



# Function Overview (Revised)

Epi Info  
Check Code  
"Code  
Industry and  
Occupation"

Epi Info uses a disconnected approach that uses machine learning and computerized "neural networks" included in the "Code Industry and Occupation" program to code entries



**Code Industry and Occupation**

Possible Matches:

Industry	Occupation
Confidence Level: 0.98 Code: 8222 Code Match: RESTAURANTS AND OTHER FOOD SERVICES	Confidence Level: 0.98 Code: 8132 Code Match: POETS AND POETESSSES, RESTAURANT SOURCE, AND C...

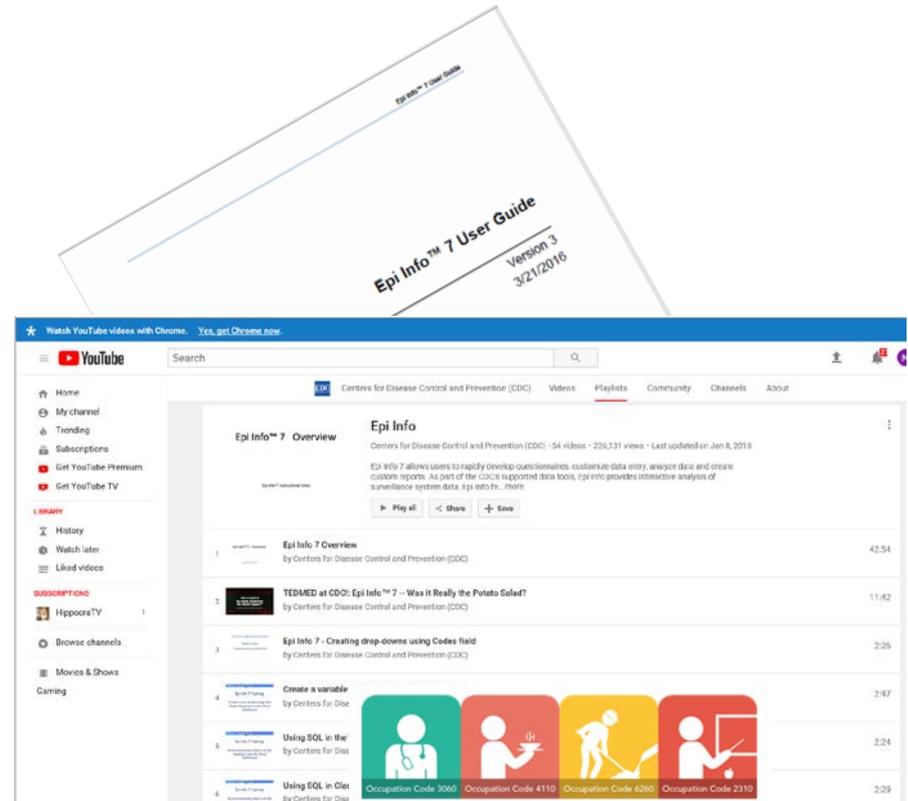
Accept

Accept

**Live Demonstration**

# Planned Project Outputs

- Epi Info
  - ✓ Check code program and template
  - ✓ User Guide entries
  - Tutorial videos
  - Scenario-based tutorials
- NIOSH
  - Topic webpage
    - Background and best practices information
      - Collecting, coding, and analyzing industry and occupation data
  - Informational/instructional webinars

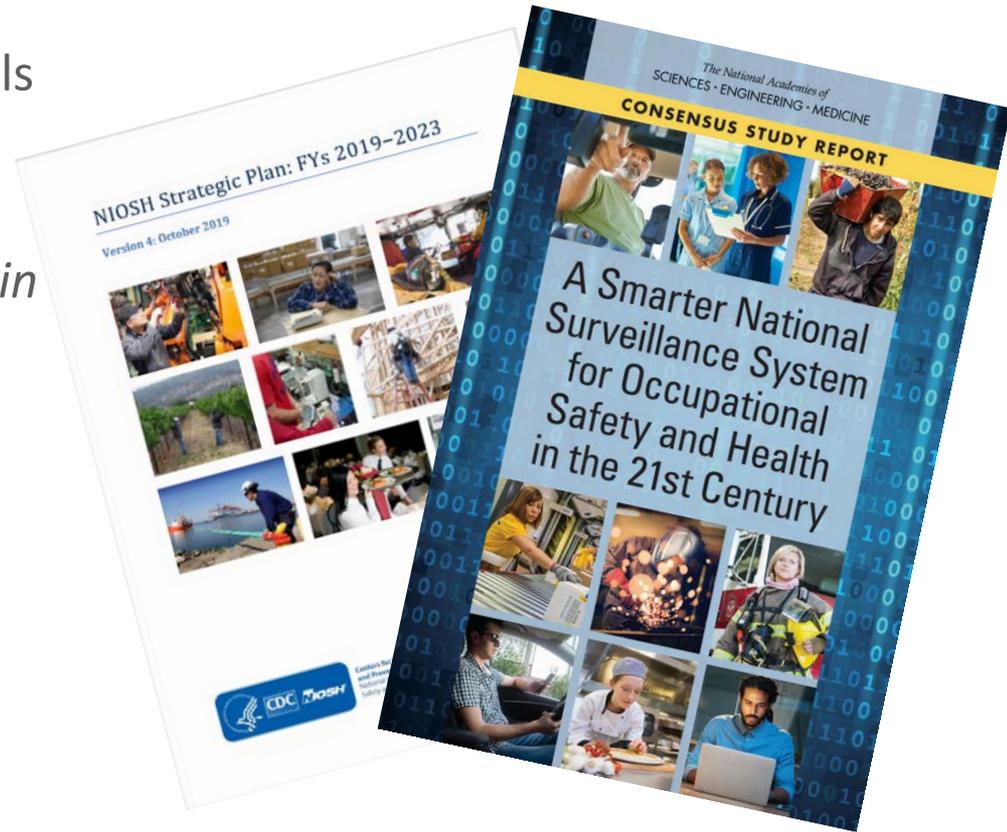


# Expected Project Outcomes

- Increased collection and use of coded industry and occupation data in public health datasets from field surveys and surveillance systems
- Increased attention to occupational health among Epi Info users
  - Emergency preparedness and response
  - Reportable disease surveillance
  - Outbreak investigation/field epidemiology
  - State and local health departments
- Improved recognition and understanding of the contribution of occupational exposures to morbidity and mortality from infectious disease and other hazards

# Looking Forward

- Advancing NIOSH's strategic goals
- Realizing *A Smarter National Surveillance System for Occupational Safety and Health in the 21st Century*
- Increased collection and use of industry and occupation data



# Questions for the Board of Scientific Counselors

- How can NIOSH best communicate the importance of collecting industry and occupation data in field surveys and surveillance systems to public health communities of practice outside of occupational health?
- How can NIOSH best promote the availability and provide instructional material for the use of these new tools to public health communities of practice outside of occupational health?
- How can NIOSH best identify and advocate for the inclusion of these new tools in additional platforms and systems used for field and surveillance data collection?

## Acknowledgements:

### CDC/NIOSH

Kelly Vanoli (NIOSH/DFSE/HIB)

Stacey Marovich (NIOSH/DFSE/HIB)

### CDC/CSELS

Sachin Agnihotri (OPHSS/CSELS/DHIS)

Asad Islam (OPHSS/CSELS/DHIS)

David Brown (OPHSS/CSELS/DHIS)

For more information, contact CDC

1-800-CDC-INFO (232-4636)

TTY: 1-888-232-6348 [www.cdc.gov](http://www.cdc.gov)

## Contact Info:

**Matthew Groenewold**

NIOSH/DFSE/HIB

(513) 458-7126

[gyr5@cdc.gov](mailto:gyr5@cdc.gov)

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

