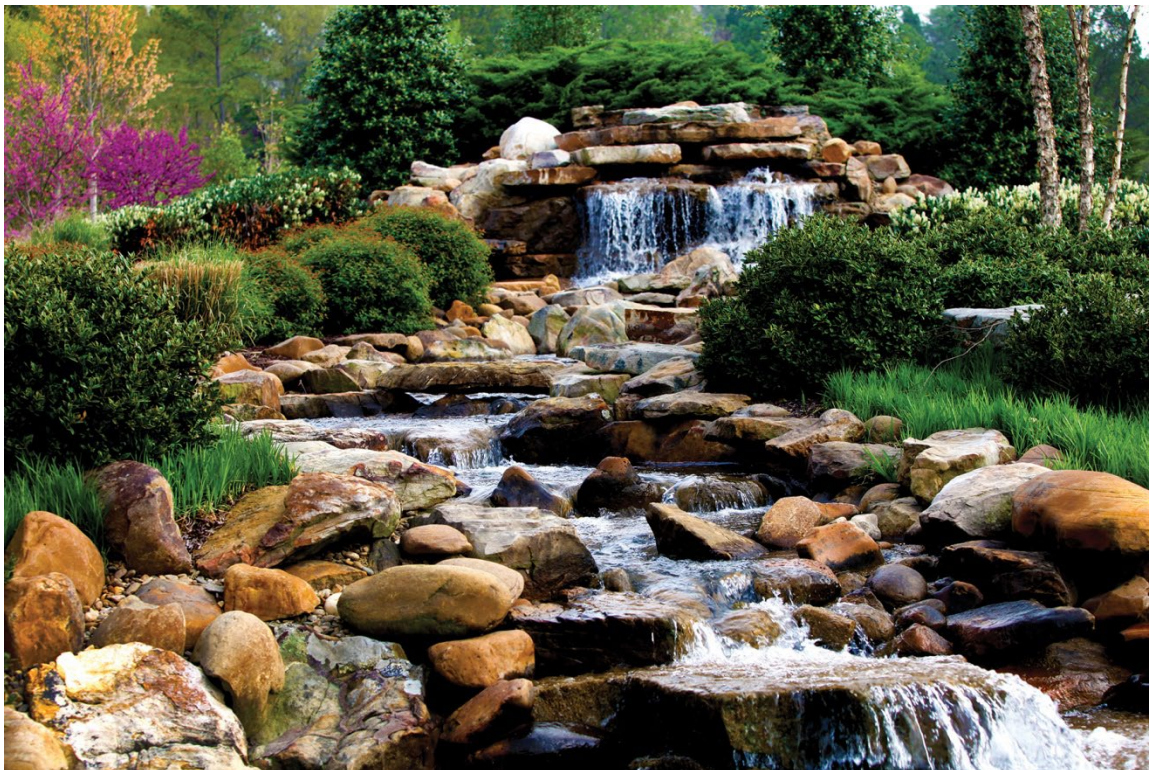


# VA EDH Data Curation Documentation FY23-Q4



Hilda B. Klasky  
Kevin Sparks  
Alina Peluso  
Aaron Myers  
Alec Hamaker  
Michael McGee  
Jonathan Zhang  
Jeremy Logan  
Heidi Hanson  
Rochelle Watson  
Anuj Kapadia

**September 2023**

## DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via OSTI.GOV.

**Website** [www.osti.gov](http://www.osti.gov)

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
**Telephone** 703-605-6000 (1-800-553-6847)  
**TDD** 703-487-4639  
**Fax** 703-605-6900  
**E-mail** [info@ntis.gov](mailto:info@ntis.gov)  
**Website** <http://classic.ntis.gov/>

Reports are available to US Department of Energy (DOE) employees, DOE contractors, Energy Technology Data Exchange representatives, and International Nuclear Information System representatives from the following source:

Office of Scientific and Technical Information  
PO Box 62  
Oak Ridge, TN 37831  
**Telephone** 865-576-8401  
**Fax** 865-576-5728  
**E-mail** [reports@osti.gov](mailto:reports@osti.gov)  
**Website** <https://www.osti.gov/>

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Computational Sciences & Engineering Division

**VA EDH DATA CURATION DOCUMENTATION FY23-Q4**

Hilda B. Klasky  
Kevin Sparks  
Alina Peluso  
Aaron Myers  
Alec Hamaker  
Michael McGee  
Jonathan Zhang  
Jeremy Logan  
Heidi Hanson  
Rochelle Watson  
Anuj Kapadia

September 2023

Prepared by  
OAK RIDGE NATIONAL LABORATORY  
Oak Ridge, TN 37831  
managed by  
UT-BATTELLE LLC  
for the  
US DEPARTMENT OF ENERGY  
under contract DE-AC05-00OR2272



# CONTENTS

CONTENTS.....	iii
1. Introduction.....	1
1.1 Background.....	1
2. Documentation Overview.....	1
2.1 Recommended Citation For FY23-Q4 Data Curation Documentation’s Sponsor Report.....	2
2.2 Previous Document Releases.....	2
3. Contents and Structure.....	5
3.1 Dataset Curation Documentation Standard Format.....	5
3.2 Dataset Conventions.....	5
3.3 Meta Table.....	6
3.4 Error Checking.....	7
3.5 FIPS.....	7
4. Internet Access Services.....	9
4.1 Data Source.....	9
4.2 Description.....	9
4.3 Inclusion.....	9
4.4 Resources.....	10
4.5 Update Frequency.....	10
5. National Instant Criminal Background Check System.....	13
5.1 Data Source.....	13
5.2 Description.....	13
5.3 Inclusion.....	13
5.4 Resources.....	14
5.5 Update Frequency.....	14
6. References.....	18
Appendix A. ERROR CHECKING.....	A-2

## 1. INTRODUCTION

The health and well-being of the Nation's men and women who have served in uniform is the highest priority for the U.S. Department of Veterans Affairs (VA). VA is committed to providing timely access to high-quality, recovery-oriented, evidence-based mental health care that anticipates and responds to Veterans' needs and supports the reintegration of returning service members into their communities. Since its creation, the VA has been working to eliminate suicide among all veterans by developing and implementing innovative suicide prevention approaches and resources.

Health outcomes, such as suicide, are typically modeled as a function of genetics and environment, where environment refers to factors beyond medical, e.g., air quality, access to transportation and food, homelessness status, etc. Mental health outcomes for each individual are considered to be associated with multiple stressors that fall under a variety of categories, including socioeconomic, economic, and physical environments. Understanding the relationships between these stressors, covariates, and health outcomes requires curated, standardized data that can be input into the VA's Recovery Engagement and Coordination for Health, Veterans Enhanced Treatment (REACH VET) or other health outcomes model.

Environmental Determinants of Health (EDH), as defined by the World Health Organization (WHO), refer to clean air, stable climate, adequate water, sanitation and hygiene, safe use of chemicals, protection from radiation, healthy and safe workplaces, sound agricultural practices, health-supportive cities and built environments, and a preserved nature, which are all prerequisites for good health.

### 1.1 BACKGROUND

With funding from the VA Office of Mental Health and Suicide Prevention (OMHSP), the EDH project has developed novel datasets associated with select health outcomes, a methodology for converting spatiotemporal data from one spatial reference (such as a 1km grid) to another (such as US Census Tracts), and health outcomes modeling capabilities. The datasets are an advancement to the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health (SDoH) covariates as key gaps are addressed. We include a finer spatial resolution (Census Tract) and environmental covariates.

The process of curating and standardizing these datasets is non-trivial, as they are often measured at different spatial and temporal resolutions and have different spatial and temporal granularities. For example, the US Census data products typically use census blocks, block groups, or counties, whereas air pollutants from the US Environmental Protection Agency (EPA) and weather data are available on 1km grids, and some economic data may be available only at a zip code level. In this context, standardized refers to the datasets all being at the same scale of spatial extent (e.g., US Census Tract and/or County), and curated refers both to a process that is repeatable, has data provenance, and which uses appropriate methodologies for converting covariates.

The data contained in the EDH datasets are drawn from multiple sources, and variables may have differing degrees of availability, patterns of missing data, and methodological considerations across sources, geographies, and years.

## 2. DOCUMENTATION OVERVIEW

This data source documentation report provides researchers with information on the structure and contents of the datasets, as well as descriptions of the data sources utilized to populate the data files. This

document covers EDH's Fiscal Year 2023, Fourth Quarter (FY23-Q4) dataset curation documentation. The datasets included are as follows:

1. Internet Access Services, 2021 and 2022, by Census tract (update).
2. National Instant Criminal Background Check System (NICS), 2021-2023, by state (update).

## **2.1 RECOMMENDED CITATION FOR FY23-Q4 DATA CURATION DOCUMENTATION'S SPONSOR REPORT**

Klasky, H.B., Sparks, K., Peluso, A., K., Myers, A., Hamaker, A., McGee, M., Zhang, J., Logan, J., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q4, ORNL/SPR-2023/3097 PUB ID 202517, 2023.

## **2.2 PREVIOUS DOCUMENT RELEASES**

Since its beginning, the following EDH project releases (datasets and data curation documentation sponsor reports) have been made available:

EDH Data Curation Documentation delivered on FY21 [1] <https://www.osti.gov/biblio/1854468>.

EDH Data Curation Documentation delivered on FY22-Q1 [2], <https://www.osti.gov/biblio/1854460>:

- Social Capital Index
  - resolution: county, 2019
  - source: ORNL
- Social Vulnerability Index
  - resolution: census tract, 2018, (only Washington D.C., Maryland, Pennsylvania, Virginia)
  - source: Centers for Disease Control, Agency for Toxic Substances and Disease Registry
  - alternate name: SVI
- Area Deprivation Index
  - resolution: block group, 2019 (only Washington D.C.)
  - source: Neighborhood Atlas, Uni of Wisconsin
  - alternate name: ADI
- Low Food Access
  - resolution: custom geometry, 2017 (only Washington D.C.)
  - source: Open Data DC

EDH Data Curation Documentation delivered on FY22-Q2 [3], <https://www.osti.gov/biblio/1862127>:

- Eviction Rates
  - resolution: county, 2000-2016 (some missing states)
  - source: Eviction Lab
- Income Inequality
  - resolution: block group, 2019
  - source: American Community Survey
- Individual-Oriented Social Vulnerability Index
  - alternate name: IOSVI
  - resolution: block group, 2019
  - source: ORNL, Census Bureau
- National Instant Criminal Background Check System
  - alternate name: NICS
  - resolution: state, 2022

- source: Federal Bureau of Investigation

EDH Data Curation Documentation delivered on FY22-Q3 [4], <https://www.osti.gov/biblio/1876283>:

- Veteran Population Status
  - resolution: county, 2020
  - source: American Community Survey
- Social Connectedness
  - resolution: county, 2021
  - source: Facebook
- Small Area Estimates of Housing Characteristics
  - resolution: block group, 2019
  - source: Census Bureau
- Internet Access Services
  - resolution: tract, 2019
  - source: Federal Communications Commission
- Medicare Part D Opioid Prescription Rates
  - resolution: county, 2019
  - source: Centers for Medicare & Medicaid Services (CMS)
- High Intensity Drug Trafficking Areas
  - alternate name: HIDTA
  - resolution: county, 2018-21
  - source: Washington/Baltimore High Intensity Drug Trafficking Areas Program

EDH Data Curation Documentation delivered on FY22-Q4 [5], <https://www.osti.gov/biblio/1892396>:

- Occupational Employment and Wage Statistics
  - alternate name: Mental Health Care Professionals per capita
  - resolution: state, 2021
  - source: Bureau of Labor Statistics
- National Survey on Drug Use and Health
  - alternate name: NSDUH
  - resolution: state, 2019
  - source: Substance Abuse and Mental Health Services Administration
- National Mental Health Services Survey
  - alternate name: N-MHSS
  - resolution: state, 2018
  - source: Substance Abuse and Mental Health Data Archive

EDH Data Curation Documentation delivered on FY23-Q1 [6], <https://www.osti.gov/biblio/1909101>:

- *State and Local Policies*
  - Naloxone laws
    - resolution: state, 2017
    - source: Rand
  - Good Samaritan laws
    - resolution: state, 2018
    - source: Rand
- Area Deprivation Index
  - resolution: block group, 2020
  - source: Uni. of Wisconsin
- Opioid Mortality Rate
  - resolution: county, 2014-2018
  - source: OEPS, Uni. of Chicago



- Opioid Prescribing Rate
  - resolution: county, 2019
  - source: OEPS, Uni. of Chicago

EDH Data Curation Documentation delivered on FY23-Q2 [7], <https://www.osti.gov/biblio/1971721>:

- Total Household Income
  - resolution: county, 2016-2021
  - source: American Community Survey
- Medicare Part D Opioid Prescription Rates (update)
  - resolution: county, 2013-2020
  - source: Centers for Medicare & Medicaid Services (CMS)
- Poverty
  - resolution: county, 2016-2021
  - source: American Community Survey
- Rural Urban Continuum Codes
  - resolution: county, 2013
  - source: Census Bureau, Department of Agriculture
- Social Capital Atlas - Civil Engagement
  - resolution: county, 2022
  - source: Social Capital Atlas
- Social Capital Atlas - Cohesiveness
  - resolution: county, 2022
  - source: Social Capital Atlas
- Social Capital Atlas - Economic Connectedness
  - resolution: county, 2022
  - source: Social Capital Atlas
- Local Unemployment
  - resolution: county, 2018-2021
  - source: Bureau of Labor Statistics

EDH Data Curation Documentation Report delivered on FY23-Q3 [8], <https://www.osti.gov/biblio/1992724>:

- Population Weighted Average Elevation
  - resolution: county, 2020
  - source: United States Geological Survey, Jim VanDerslice (Associate Professor in the Division of Public Health at University of Utah)
- Education Attainment
  - resolution: county, 2016-2021
  - source: US Census Bureau, American Community Survey
- Eviction Rates (update)
  - resolution: county, 2016-2021
  - source: The Eviction Lab (Princeton University)
- Food Insecurity
  - resolution: county, 2010-2021
  - source: Feeding America, US Hunger Relief Organization

### 3. CONTENTS AND STRUCTURE

#### 3.1 DATASET CURATION DOCUMENTATION STANDARD FORMAT

Each data source description follows a standard format with the following fields:

- Source (name of the organization that provided the raw data, e.g., Health Resources and Services Administration [HRSA] for the Area Health Resources Files [AHRF]). Prior to the FY23Q4 release, we used the name “sponsor” for the source organization that provided the raw data.
- Description (brief, general description of the data)
  - Inclusion in the EDH datasets
    - Lists the social or environmental determinants of health domains to which the data source has contributed variables.
    - Includes additional information about the data source relevant to the EDH dataset.
- Resources (links to original data source documentation, data download sites, and other relevant information).
- Update frequency: how often is each dataset going to be updated.
- Variable definitions and specifications (in tabular format):
  - Variable name (column name)
  - Variable label (optional, if different than the variable or column name)
  - Source table, if multiple data tables were available from the original data source (optional)
  - Numerator (for derived variables; optional)
  - Denominator (for derived variables) or original variable (when renamed for the EDH dataset, optional)
  - total\_rows: how many rows are in each column in each dataset (Starting in FY23Q2 )
  - null\_rows: for each column in each dataset, how many rows are null. (Starting in FY23Q2 )
- Variable availability across years (in tabular format):
  - Variable name (column name)
  - Data year availability (e.g., 2009 to 2018)

#### 3.2 DATASET CONVENTIONS

Regarding datasets’ versioning, the Microsoft SQL Server database system is used to supply the datasets. A table is used to hold each dataset. The following schema names for the quarterly releases are included in the database: OMHSP\_FY22Q1, OMHSP\_FY22Q2, OMHSP\_FY22Q3, OMHSP\_FY22Q4, OMHSP\_FY23Q1, ..., and so on. These will aid in distinguishing between releases when we deliver the same data set, updated, from one release to another.

Variables in the EDH dataset were created from these several data sources in one of two ways:

1. Drawn directly from the original data source. When the data was available from the data source as needed, we renamed the original variables for clarity and consistency across years, and to fit the naming conventions of the SEDH data files.
2. Derived using data from the original data source. For some data sources, it was necessary to calculate percentages or rates for inclusion in the data files. The numerators and denominators for the variables and their sources are shown following each data source description.

The following conventions were followed in constructing the SEDH datasets to provide researchers with a consistent and easy-to-use resource:

- Variable assignment to annual datasets. Variables appear in the annual datasets that correspond with (1) the single year represented by the original data source (e.g., US Area Deprivation Index 2020), or (2) the last year in a period represented by the data (e.g., American Community Survey data aggregated over 2012 to 2016 is in the 2016 dataset).
- Variable availability. The availability of each variable changes across data years. Following each data source description in this report is a table showing the availability of each variable in the annual datasets. When a variable is not available, we indicate it with NA (not available) or simply ‘-’.
- Variable naming. Except for the geographic ID variables, all variable names begin with a data source acronym followed by an underscore and a descriptive title.
- Missing values. The datasets use a blank to denote a missing value, almost exclusively. The one exception is the provider ratio variables from the County Health Rankings (CHR) data, which have negative values for counties where the number of providers is zero. This is described further in the description of the CHR data.

Detailed information about each data source is included in the following sections of this report.

### 3.3 META TABLE

Starting with release FY23Q1, the ORNL team has been supplying an updated metadata table called SEDH\_meta\_table, which is located in the OMHSP schema. SEDH stands for Social and Environmental Determinants of Health repository. The columns of the SEDH\_meta\_table are as follows:

- schema. The schema names for the quarterly releases are included in the database: OMHSP\_FY22Q4, OMHSP\_FY23Q1, OMHSP\_FY23Q2, ..., and so on;
- table\_name. The table name as it appears in the MS SQL Server database;
- table\_name\_description. The description of the table name;
- column\_name. Each of the dataset’s column names as they appear in the MS SQL Server table;
- column\_name\_description. The description of the column name;
- availability\_across\_years. The years for which the data is available;
- reference\_report. The reference to the ORNL’s report that has this dataset’s data curation documentation;
- report\_url. The URL link to the ORNL report;
- column\_type. The column type as it appears in the MS SQL Server table;
- column\_length. The column length as it appears in the MS SQL Server table;
- total\_rows: how many rows are in each column in each dataset (starting in FY23Q2)
- null\_rows: for each column in each dataset, how many rows are null (starting in FY23Q2)
- data\_source. The name of the source organization that provided the raw data (starting in FY23Q4)
- data\_source\_description. The description of the source organization (starting in FY23Q4)
- data\_source\_url. The URL of the source organization (starting in FY23Q4)
- data\_categories. General data categories such as social, economic, educational, etc. (starting in FY23Q4)
- spatial\_resolution. The spatial resolution or geography, for example: state, county, block group, tract, etc. (starting in FY23Q4)

With each new quarterly release, the metadata table will be updated with new details in the aforementioned columns for each delivered dataset.

Please take note that the report\_url column with the URL link will be updated in the VA’s CDW transmit database as soon as it becomes accessible on the Office of Scientific and Technical Information website (osti.gov) of the US Department of Energy, which is typically four weeks after each quarterly release.

### 3.4 ERROR CHECKING

Beginning with the FY23Q1 release, the ORNL team will additionally give succinct information regarding error checking activities in order to provide formal evidence that the datasets supplied have been thoroughly error checked. Our data profiling process is described in our project’s overview manuscript [9]:

“Following standard data and software development methodologies, data profiling is performed in four different work environments: 1) a team-shared work environment for selection, extraction, and refinement of raw data (development); 2) an ORNL intranet work environment focused on quality assurance testing (QA-Intra); 3) an ORNL Knowledge Discovery Infrastructure (KDI) secure work environment that stores highly sensitive data and ensures its security (QA-KDI). And finally, 4) a production environment housed within the KDI environment and accessible to our VA sponsors, (Production). We carried out test iterations in each of the four work environments as the datasets moved through them to confirm data integrity and system compatibility.

All datasets were error-checked using a data profiling strategy that includes at least two reviewers and the following test groups:

3. evaluating missingness: i.e. determining the amount of missing data by randomly checking for them;
4. compiling descriptive statistics, such as the number of rows, columns, and types of variable data;
5. appending checksums to a subset of the columns on both the source and destination copies to ensure consistency;
6. consistently representing the social and physical environment using FIPS codes as geographic administrative boundaries and confirming that the FIPS codes correspond to the geographic administrative boundaries of the original data;
7. manually comparing the first, last, and five additional randomly selected rows for consistency between the source and target datasets.

When datasets are developed at ORNL, which we call ‘derivative’, ORNL will provide extra error-checking utilizing a combination of statistical methodologies based on each dataset’s properties, in addition to the data profiling methodology described above.” [9]

The error-checking results for FY23Q4 follows:

Dataset Name	Rows	Columns	Development		QA-Intra		QA-KDI		Production (Transmit)		Error ratio
			Passes	Fails	Passes2	Fails3	Passes4	Fails5	Passes6	Fails7	
OMHSP_FY23Q4 - fcc_internet_access_tract_jun_2020	73767	3	5	0	5	0	5	0	5	0	0
OMHSP_FY23Q4 - fcc_internet_access_tract_jun_2021	73767	3	5	0	5	0	5	0	5	0	0
OMHSP_FY23Q4 - nics_firearm_background_checks_state_2021_2023	1728	6	3	2	3	2	4	1	5	0	0.33
OMHSP.SEDH_meta_table	740	17	3	2	3	2	4	1	5	0	0.33

More information about error checking results can be found in Appendix A.

### 3.5 FIPS

ORNL employs the Federal Information Processing Standards (FIPS) as geographic identifiers and, in database terms, as the primary key in each dataset or table for the purposes of this project. The FIPS of the United States are a set of publicly announced standards that the National Institute of Standards and

Technology (NIST) developed for use in computer systems and non-military applications. Particularly, the US government has developed various FIPS specifications to standardize codes for geographical areas in which FIPS is a unique identifier associated with a region of space or geographical area. The specifications for different geographical areas are as follows: FIPS 10-4 is for country codes or region codes, FIPS 5-2 is for state codes, and FIPS 6-4 is for counties.

Smaller geographic organizations' FIPS codes are generally unique inside bigger geographic entities. FIPS state codes, for example, are unique within a country, while FIPS county codes are unique within a state. Since counties nest inside states, a full county FIPS code specifies both the state and the nesting county. For example, there are 49 counties in the 50 states that finish with the number "001". To make these county FIPS codes distinct, the state FIPS codes are appended to the front of each county (01001, 02001, 04001, etc.), where the first two numbers relate to the state the county is in, and the final three digits correspond to the county explicitly.

The use of FIPS in the US has continued, despite NIST's gradual withdrawal initiated in 2002. NIST intended to replace FIPS with the Geographical Name Information System (GNIS) Feature ID in early 2002. Census.gov defines a geographic identifier, or GEOID, as a "unique identifier to geographic entities that facilitates organization and presentation." GEOIDs are maintained by the US Board on Geographic Names, however, after 10 years of replacement efforts, in 2010, and until now, many federal organizations in the United States were still using FIPS. For example, the US Census Bureau employed FIPS codes to identify legal and statistical entities for counties, because neither FIPS codes nor GNIS codes provide appropriate coverage of many smaller geographic areas due to the fuzziness and vagueness of natural boundaries. As a matter of fact, the US Census Bureau creates and maintains codes that include census divisions, regions, tracts, block groups, census blocks, and urban areas. A depiction of the hierarchy is found at the US Census Bureau's standard hierarchy of census geographic entities, which is found here <https://www2.census.gov/geo/pdfs/reference/geodiagram.pdf>

As the main key for all the datasets created for this project, we have followed the convention of utilizing the column "FIPS" as the column name to uniquely identify the data, regardless of the source FIPS granularity. We indicate the FIPS granularity, such as region, state, county, census division, tracts, group blocks, etc. in the metadata table and reports' descriptions. We presume that users of these datasets are familiar with joining several datasets using FIPS columns with different geographic area levels.

We have only provided one dataset with no FIPS column since the beginning of this project, as follows:

1. The National Mental Health Services Survey (table: national\_mental\_health\_services\_survey) with raw data received from the source was delivered in FY22Q4.

The datasets (tables) above were a special request from the sponsor.

## 4. INTERNET ACCESS SERVICES

### 4.1 DATA SOURCE

Federal Communications Commission (FCC)

### 4.2 DESCRIPTION

This dataset is an update to the FY22Q3's internet access services previously delivered, in which we provided the status of internet access services as of June 30, 2019. In the FY23Q4 delivery we are providing FCC's internet access service data for June 30, 2020 and June 30, 2021 in two separate tables i.e., [OMHSP\_FY23Q4].[fcc\_internet\_access\_tract\_jun\_2020] and [OMHSP\_FY23Q4].[fcc\_internet\_access\_tract\_jun\_2021].

This dataset provides data on total and residential Internet access connections by downstream and upstream speed, by technology, by geography, and over time. Internet access connections are those in service at over 200 kilobits per second (kbps) in at least one direction and reported to the FCC through Form 477. This dataset provides data from residential fixed connections per 1000 households by census tract (N = 73,767).

The description of the code per connections follows:

Code	Connections per 1,000 Households
0	Zero
1	Zero < x <= 200
2	200 < x <=400
3	400 < x <=600
4	600 < x <=800
5	800 < x

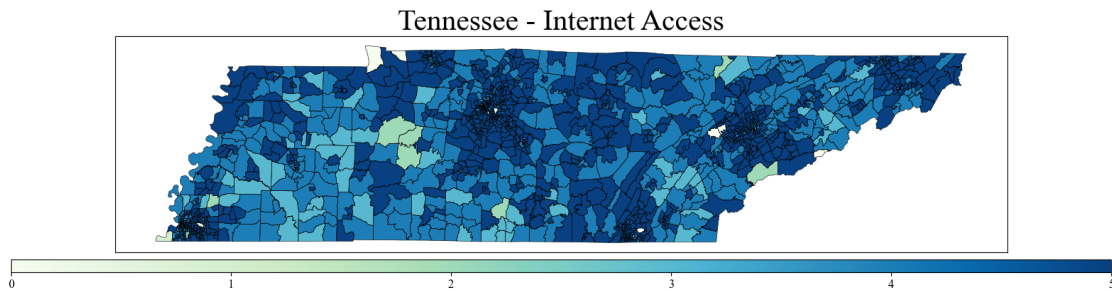


Figure 1. Internet Access by Census tract as of 2021 for the state of Tennessee.

### 4.3 INCLUSION

Year: 2020, 2021.

Geographical unit: Census tract. Continental US, plus the following incorporated organized territories: Guam, Puerto Rico, and Virgin Islands.

#### **4.4 RESOURCES**

For more information on Internet Access Services:

Main: <https://www.fcc.gov/internet-access-services-reports>

#### **4.5 UPDATE FREQUENCY**

This dataset will be updated and distributed every fiscal year, or as requested by the sponsor. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

**Table 1 . Internet Access Services ( INTERNET\_ACCESS )**

<b>variable name</b>	<b>variable label</b>
FIPS	Federal Information Processing Standards (FIPS) at census tract level.
pcat_all	Residential Fixed High-Speed Connections over 200 kbps in at least one direction per 1000 Households.
pcat_10x1	Residential Fixed High-Speed Connections at least 10 Mbps downstream and at least 1 Mbps upstream per 1000 Households.



**Table 2 . Variable availability across years, ( INTERNET\_ACCESS )**

<b>variable name</b>	<b>2020</b>	<b>2021</b>	<b>total rows</b>	<b>null rows</b>
FIPS	X	X	73767	0
pcat_all	X	X	73767	0
pcat_10x1	X	X	73767	0

## 5. NATIONAL INSTANT CRIMINAL BACKGROUND CHECK SYSTEM

### 5.1 DATA SOURCE

Federal Bureau of Investigation

### 5.2 DESCRIPTION

This dataset is an update to the NICS delivered on the FY22Q2 release, where we provided an update on the status of NICS from January 1998 to June 2021. This dataset provides the number of permits and firearm transactions from 2021-01 to 2023-08 by state.

As required by law, the NICS conducts background checks on people who want to own a firearm or explosive. When a buyer attempts to purchase a firearm, the seller, known as a Federal Firearms Licensee (FFL), contacts NICS either electronically or by phone. The prospective buyer completes the ATF form, which is forwarded to the NICS by the FFL. The buyer is subjected to a background check by the NICS staff. This background check ensures the buyer has no criminal history and is not otherwise ineligible to purchase or own a firearm. Since its inception in 1998, more than 300 million checks have been performed, resulting in over 1.5 million denials.

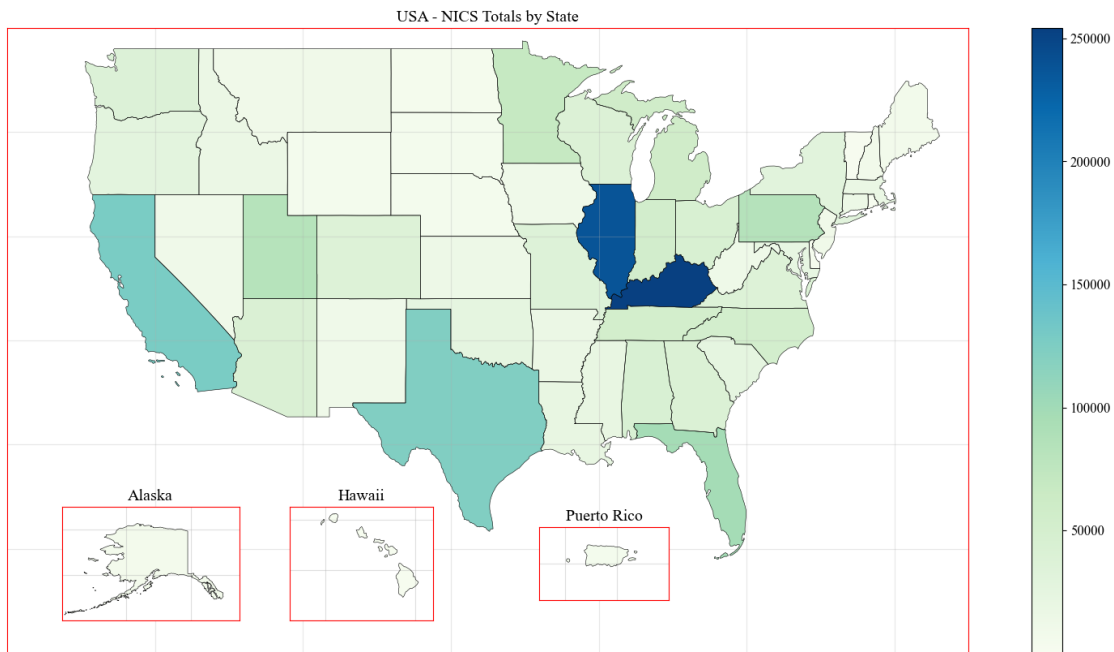


Figure 2. NICS Totals by State as of August 2023.

We renamed the original column [month] as year-month.

### 5.3 INCLUSION

Year: from 2021-01 to 2023-08

Geographical unit: FIPS at State Level, Continental US plus the following incorporated organized territories: Guam, Puerto Rico, and Virgin Islands.

## 5.4 RESOURCES

National Instant Criminal Background Check System (NICS):

<https://raw.githubusercontent.com/BuzzFeedNews/nics-firearm-background-checks/master/data/nics-firearm-background-checks.csv>

<https://www.fbi.gov/services/cjis/nics>

## 5.5 UPDATE FREQUENCY

This dataset will be updated and distributed every fiscal year, or as requested by the sponsor. Minimal quarterly updates may be necessary to correct minor data inaccuracies.

**Table 3 . National Instant Criminal Background Check System ( NICS )**

<b>variable name</b>	<b>variable label</b>
FIPS	State level FIPS.
state	State of residence of persons on which background checks are conducted.
permit	Number of permits approved.
handgun	Number of handgun transactions.
totals	Total number of firearm transactions, including handguns, long guns, private sales, returns to seller, firearm rentals.
year_month	The year and month of data collection.

**Table 4 . Variable availability across years, ( NICS )**

<b>variable name</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>total rows</b>	<b>null rows</b>
FIPS	X	X	X	1728	0
state	X	X	X	1728	0
year	X	X	X	1728	0
permit	X	X	X	1728	0
handguns	X	X	X	1728	0
totals	X	X	X	1728	0



## 6. REFERENCES

- [1] Christian, J.B., Branstetter, M, Klasky, H.B., Tuccillo, J., Sparks, K., Rastogi, D., Watson, R., Yoon, H-J., Kim, Y., VA EDH Data Curation Documentation - FY 2021, Rev. 2, ORNL/SPR-2021/2366 - Pub ID 170648. 2021. <https://www.osti.gov/biblio/1854468>
- [2] Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Devineni, P., and Watson, R. VA EDH Data Curation Documentation - FY22-Q1, Rev. 2, ORNL/SPR-2022/2316- Pub ID 172755. 2022. <https://www.osti.gov/biblio/1854460>
- [3] Christian, J.B., Klasky, H.B., Sparks, K., Peluso, A., Tuccillo, J., Rastogi, D., Branstetter, M., Whitehead, M., Hamaker, A., and Watson, R., VA EDH Data Curation Documentation - FY22-Q2, Rev. 2, ORNL/SPR-2022/2391 - Pub ID 174092. 2022. <https://www.osti.gov/biblio/1862127>
- [4] Klasky, H.B., Sparks, K., Logan, J., Tuccillo, J., Whitehead, M., Hamaker, A., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q3, Rev. 2. ORNL/SPR-2022/2487 - Pub ID 178645. 2022. <https://www.osti.gov/biblio/1876283>
- [5] Klasky, H.B., Sparks, K., Logan, J., Hamaker, A., Whitehead, M., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY22-Q4, ORNL/SPR-2022/2587, PUB ID 183700. 2022. <https://www.osti.gov/biblio/1892396>
- [6] Klasky, H.B., Sparks, K., Logan, J., Hamaker, A., Whitehead, M., Peluso, A., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q1, ORNL/SPR-2022/2694, PUB ID 187842. 2022. <https://www.osti.gov/biblio/1909101>
- [7] Klasky, H.B., Sparks, K., Peluso, A., Whitehead, M., K., Logan, J., Hamaker, A., McGee, M., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q2, ORNL/SPR-2023/2857, PUB ID 19179. 2023. <https://www.osti.gov/biblio/1971721>
- [8] Klasky, H.B., Sparks, K., Peluso, A., K., Logan, J., Hamaker, A., McGee, M., VanDerslice, J., Hanson, H., Watson, R., and Kapadia, A., VA EDH Data Curation Documentation - FY23-Q3, ORNL/SPR-2023/2930 PUB ID 195499, 2023. <https://www.osti.gov/biblio/1992724>
- [9] Klasky, H.B., Hanson, H., Sparks, K., Whitehead, M., Blair, C., and Kapadia, A., “Dataset Repository for Investigating Suicide Risk Using Social and Environmental Determinants of Health”, ORNL/TM-2023/3027 Pub ID 183902. 2022. <https://www.osti.gov/biblio/1997699>

## **APPENDIX A. ERROR CHECKING**



**APPENDIX A. ERROR CHECKING**

This section lists descriptive statistics of the datasets provided for the FY23Q4 delivery. The following statistics were run using Python’s describe() function.

**Table A1 – [ OMHSP\_FY23Q4].[fcc\_internet\_access\_tract\_jun\_2020]**

	fips (tractcode)	pcat_all	pcat_10x1
count	73767	73767	73767
mean	28438652274	4.5392113	4.28964171
std	16524358024	0.801304512	0.9801843
min	1001020100	0	0
25%	13045910651	4	4
50%	28049010600	5	5
75%	42003465700	5	5
max	78030961200	5	5

**Table A2. – [ OMHSP\_FY23Q4].[fcc\_internet\_access\_tract\_jun\_2021]**

	fips (tractcode)	pcat_all	pcat_10x1
count	73767	73767	73767
mean	28438652274	4.639757615	4.44374856
std	16524358024	0.744280388	0.895514271
min	1001020100	0	0
25%	13045910651	4	4
50%	28049010600	5	5
75%	42003465700	5	5
max	78030961200	5	5

**Table 3. [OMHSP\_FY23Q4].[nics\_firearm\_background\_checks\_state\_2021\_2023]**

	fips	permit	handgun	totals
count	1728	1728	1728	1728
mean	31.35185185	12698.31655	14005.95718	51302.76215
std	18.18864784	41863.53595	15277.96775	90226.8622
min	1	0	0	79
25%	17	742	3567.75	9516
50%	30.5	3351.5	8491.5	25589.5
75%	45	13018.75	20092	58885.5
max	78	480752	97406	1427917