Bubbles (Bethany) Cram

03/18/2023

LIS 545B

**Term Project – Final Report**

**Data Set from Term Project Data & Metadata Profile (Note, Data.gov moved data):** <https://catalog.data.gov/dataset/state-summaries-washington>

**Repository Profile Data:**

<https://www.datafiles.samhsa.gov/dataset/national-mental-health-services-survey-2015-n-mhss-2015-ds0001>

**Data & Metadate Profile:**

Data informs many choices that are made within businesses and even within local and state government agencies. Interpreting data is important because of the impact it can have on many people who are affected by decisions. The data I have chosen to review is the VA State Summary Report, then narrowed that down to the “Washington FY15 Summary”. The data is from the “National Center for Veterans Analysis and Statistics” (Department of Veterans Affairs, 2022). There is also data from the Census Bureau, the Department of Defense, and Integrated Public Use Microdata Series (IPUMS) for multiple years and multiple states and territories.

There are multiple stakeholders involved in this data set, the veterans themselves, the agencies involved, the employees of the agencies, the US Government, elected officials, and taxpayers. There are approximately 1,281 data sets across 65 pages, with 20 individual files per page for about 1,281 – 1,300 files in total for all summaries. Some files are in multiple file formats, formats are ZIP, HTML, PDF, CSV, RDF, JSON, and XML. The Washington FY15 summary is in PDF format. The data sets state that they are for public use and access. There are no noted restrictions over what the data.gov site lists.

No special software is required to open a ZIP file as it is possible to “extract all” files inside a ZIP folder, from there the files may need additional software to view and analyze. HTML is best to run on a web browser to see what the program looks like as the web browsers are designed to read and render HTML files for viewing and analyzing. Additionally, any text editor can be used to read and analyze a HTML file. PDFs are commonly opened and viewed with Adobe Acrobat, or on a windows computer it is possible to use the browser which acts like a PDF reader if you wish to view a PDF without special software. CVS files can be viewed with special software such as Microsoft Excel, CVS FileView, LibreOffice Calc, and numerous other software programs for viewing and analyzing. RDF files can be viewed and analyzed with CGRTools, Chemaxon Instant JChem, RDF Grapher, RDF Viewer, and other software. Viewing and analyzing a .JSON file can be done via an online JSON viewer. XML files can be read and analyzed using a text editor.

The metadata for this data set is in separate files and some files also have metadata in them. The large metadata file appears to contain approximately 20 pages of metadata identifiers for the various files within the data set. Meta data types include; keyword, contactpoint, publisher, identifier, description, title, ispartof, programcode, distribution, dataqualoty, license, accrualperiodicity, language, resource type, metadata created date, metadata updated date, maintainer, data first published, data last modified, public access level, metadata context, metadata catalog ID, schema version, catalog describedby, harvest object id, harvest source id, harvest source title, homepage url, license, source hash, and source schema version to list a few. The metadata for Washington FY15 includes the metadata fields; last updated, created, name, format, license, created, media type, id, metadata modified, package id, and state.

The list of metadata provided is extremely comprehensive for the entire umbrella of summaries and it includes all of the categories I could think of and even ones I did not think of. The metadata categories for Washington FY15 list the fields “last updated” and “created”, however the fields are filled with “unknown”. This should be kept current to improve unfamiliar users’ ability to use the data set within the repository for future use. According to the SCOPE MetaData Team (2020), data.gov hosts a catalog of government data, the catalog is managed through a CKAN system but there is no standard way that metadata is structured at this time. Some of the files have various structures to the metadata organization. This author would not say that the metadata is structured according to any metadata standard. Specifically for the Washington FY15 the metadata doesn’t appear to have been organized according to a standard structure.

The current data and meta data seem to be very well organized by types, locations, age, conflicts served, disabilities, mental health needs, and many other fields to sort by in the large data set. There are not as many types of meta data listed for Washington FY15. Overall, this author is impressed with the amount of information covered in this data set. The Department of Veteran’s affairs website has an article and publication repository written based off data and meta data; however, it is unclear if any of those articles or publications were based off the Washington FY15 summary or a combination of data sets. Publications listed within Washington FY15 are the VA Veteran Population Projection Model, VA Geographic Distribution of Expenditures, VA Annual Benefits Report, U.S. Census Bureau, and American Community Survey. This author used the Department of Veteran’s Affairs website, google, bing, msn, yahoo, and found that there was not an easily identified correlation/citation between the Washington FY 15 data set and the seemingly related articles.

**Update: New Link to Data:** <https://catalog.data.gov/dataset/state-summary-washington-fy15/resource/df339ebb-bea6-47f0-a9b1-bbaf2b527e20>

**Term Project – Repository Profile**

The data set I chose to review in the *Data and Metadata Profile* was the VA State Summary report (FY-2015), which contained information about military veterans and agencies as well as other data relating to the demographics of the veterans and health care. For the purpose of this repository profile, I chose the National Mental Health Services Survey (N-MHSS-2015) as an appropriate repository for the data set because both the data sets are from the year 2015 and both contain information about veterans, demographics, and the mental health care of this population (veterans) and it is publicly accessible. This data set was found through re3data.org and then searching for “Washington State Healthcare” then locating HealthData.gov and going to the National Mental Health Services Survey and from there viewing the “Access this Data” field.

Additionally, the National Mental Health Services Survey states that their data set is the “only source of national and state-level data on mental health service delivery system reported by both public and private specialty mental health treatment facilities…” (*National Mental Health Services Survey 2015 (N-MHSS-2015-DS0001) | SAMHDA*, n.d.) This type of data is a niche and there is not a plethora of data sets and repositories that exist within this scope of work without submitting lengthy requests for viewing potentially identifiable protected data at an in-person location at a National Mental Health office.

This repository has a defined collection scope of the United States and jurisdictions, including treatment facilities across the United States and accompanying jurisdictions. Psychiatric hospitals, general hospitals, medical facilities operated by the Department of Veterans Affairs (VA), partial hospitalization/day treatment mental health facilities, outpatient mental health facilities, residential treatment centers (RTCs) for children and adults, multi-setting mental health facilities, community mental health centers (CMHCs), and other types of residential treatment facilities were included in this scope of work and were invited to contribute data through a survey from the National Mental Health Services (NMHS). The survey sent to these facilities was voluntary so the data may be limited due to non-response. The data is not adjusted to account for non-responsive facilities. (*National Mental Health Services Survey 2015 (N-MHSS-2015-DS0001) | SAMHDA*, n.d.) The repository does not appear to be open for data submissions from the general public as it has an explicit defined collection scope.

The data comes from the above-mentioned sources and is used to compliment information obtained “through SAMHSA’s National Survey of Substance Abuse Treatment Services (N-SSATS). Treatment facility information from N-MHSS is used to populate the mental health component of SAMHSA’s Behavioral Health Treatment Services Locator” (*National Mental Health Services Survey 2015 (N-MHSS-2015-DS0001) | SAMHDA*, n.d.).

N-MHSS is an annual survey to collect information on the location, organization, structure, services, and utilization of mental health treatment facilities. It is designed to collect statistical information on all known mental health treatment facilities in the United States. This includes the 50 states, the District of Columbia, and all U.S. territories. Every other year, starting in 2014, N-MHSS also collects information on the numbers and demographics of persons served in these treatment centers. (*National Mental Health Services Survey*, n.d.)

In the VA State Summary Report FY 2015, we examined the data set specifically from Washington State about military veterans, and the National Mental Health Services Survey repository contains information about the veterans’ groups specified in the VA State Summary Report, which makes this data set an excellent choice to include the data set from the VA State Summary Report FY 2015.

It is not required to create an account to log in and download the data. On the National Mental Health Services website, a user can choose from four options to directly download the data set via a file. The file choices are SAS, SPSS, Stata, Delimited, and R. The dataset has an accompanying codebook and a copy of the questionnaire in PDF format. This author did not observe any methods to access the data set other than direct download. There are twenty-two published articles that cite data from the N-NMHSS Surveys from 2014-2022. The National Mental Health Services Survey is an annual survey; however, this repository profile is focused only on the year 2015 as the VA State Summary Report was also from 2015. This author did not note the use of a specific standardized metadata method of display.

The owner of the dataset is the Health and Human Services Officer of the Chief Data Officer and while the terminology Dissemination Information Package (DIP) is not directly used, the website with the direct data download links has a contact name of Brooklyn Lupari, SAMHSA/CBHSQ c/o SAMHDA Support and an email address of [samhda-support@samhsa.hhs.gov](mailto:samhda-support@samhsa.hhs.gov) Upon browsing the parent site, Substance Abuse and Mental Health Services Administration, this author located the following:

The Research Data Center (RDC) provides researchers with the opportunity to apply for access to restricted data files that are not available anywhere else. For access to restricted use data, please submit an application to SAMHSA RDC at rdca@samhsa.hhs.gov. A proposal review will occur once an application is submitted. The Restricted-Use Data Analysis System (RDAS) is an online tool that lets users access restricted use NSDUH files (RUFs) within the limits of predetermined variables and crosstabs. RDAS includes built in confidentiality restraints. (SAMHSA, n.d.)

There is also information which requests copies of any articles published or under consideration for publication using the data contained within this data set.

**Additional Information**

**Recommended Data Citation**

The FY 15 Summary by State from the VA recommended data citation is:

National Center for Veterans Analysis and Statistics, Contact: www.va.gov/vetdata

Sources: VA Veteran Population Projection Model, VA Geographic Distribution of Expenditures,

VA Annual Benefits Report, U.S. Census Bureau, American Community Survey

The SAMHSA and NMHSS recommended data citation is:

Data users agree to reference the following citation in any of the publications that use these public use data.

Substance Abuse and Mental Health Services Administration, National Mental Health Services Survey (N-MHSS): 2014. Data on Mental Health Treatment Facilities. BHSIS Series S-87, HHS Publication No. (SMA) 16-5000. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2016.

Authors of publications that use these data are required to send citations of their published works to SAMHSA for inclusion in a database of related publications ([SAMHSA\_support@smdi.com](mailto:SAMHSA_support@smdi.com)).

**Long-Term Preservation Considerations**

This dataset is downloadable via multiple different files such as a CSV,. JSON, PDF, or .XML file, which can be opened with many tools like Numbers (Mac-based OS), Excel, Google Sheets, a browser, etc. Depending on which file a user is opening, there are not immediate risks to having an obsolete file. There is a possibility that new software could replace these file types and then the owners of the files would need to determine if replacing or converting the files to the new software is best for long-term preservation. This seems rather unlikely as the data is already available in multiple formats for public use.

**Copyright License**

In the “Information Description” PDF from the N-MHSS file, there is a note under access and availability which states:

Graphical user interface, text, application, letter

Description automatically generated

There is no such restriction placed on the FY 15 Summary by State from the VA, however it does say that the data is for public use.

**Human Subject Considerations**

The data does not have any personally identifiable data about individual persons, however there are data available upon request to view in person that may contain personally identifiable data. This author did not submit a request to view the restricted data on site. The availability of restricted data requires an application process which leads this author to conclude that steps were taken to anonymize the data for privacy or ethical concerns prior to releasing the data for public use. This author also concludes that there may be personally identifiable data within the restricted data sets, however it is not possible to know for certain without applying to view the restricted data.

**References:**

Ball, A., Chen, S., & Alderete, K. A. (n.d.). *Metadata Standards Directory*. Retrieved January 24, 2023, from <http://rd-alliance.github.io/metadata-directory/>

*Best Practices Index*. (n.d.). DataONE. Retrieved January 24, 2023, from <https://dataoneorg.github.io/Education/bestpractices/>

Department of Veterans Affairs. (2022). *State Summaries\_Washington* [Data set]. Department of Veterans Affairs. <https://catalog.data.gov/dataset/state-summaries-washington>

*Mission & Purpose | IPUMS*. (2021). <https://www.ipums.org/mission-purpose>

*National Mental Health Services Survey*. (n.d.). <https://www.samhsa.gov/data/data-we-collect/n-mhss-national-mental-health-services-survey>

*National Mental Health Services Survey 2015 (N-MHSS-2015-DS0001) | SAMHDA*. (n.d.). <https://www.datafiles.samhsa.gov/dataset/national-mental-health-services-survey-2015-n-mhss-2015-ds0001>

SAMHSA. (n.d.). *Data We Collect*. (n.d), from <https://www.samhsa.gov/data/data-we-collect>

SCOPE Metadata Team. (2020). *Metadata Systems for the U.S. Statistical Agencies, in Plain Language*.

*State Summary: Washington FY15 - Washington - Catalog*. (n.d.). Retrieved January 27, 2023, from <https://catalog.data.gov/dataset/state-summary-washington-fy15/resource/df339ebb-bea6-47f0-a9b1-bbaf2b527e20>

U.S. Department of Veterans Affairs. (2016). *State Summaries Washington*. U.S. Department of Veteran’s Affairs. <https://www.va.gov/vetdata/docs/SpecialReports/State_Summaries_Washington.pdf>