



Enterprise Data Dissemination Informatics Exchange (EDDIE)

<https://www.phaim.health.pa.gov/EDD>

Category:

Open Government and Data, Information, and Knowledge Management

Contact:

Patrick Keating | Chief Information Officer
Pennsylvania Department of Health
2150 Herr Street, Harrisburg, PA 17103
Phone: 717.787.7521 | Email: pakeating@pa.gov

Project Initiation Date:

January 2014

Project Completion Date:

December 2015

Executive Summary:

There is more health-related data being collected and available today than any other time in history. This has resulted in more and better information to drive public health policies and program decisions.

As the volume of data grows, so too does the need for an efficient way for a wide variety of users to efficiently access and analyze the data. In the era of big data, the demand for high quality health data and statistics never seems to be satisfied. Meanwhile, the open data movement has resulted in increased expectations for data transparency and accessibility.

In response to these trends and in alignment with Governor Wolf's goal for "Government that Works" through greater collaboration with external entities and citizen engagement, the Pennsylvania Department of Health is striving to become a more data-driven organization providing useful information to the public.

The Department of Health maintains numerous applications, data formats, databases, reports, and webpages to handle the demands of health data users. In 2014, the department began a collaborative project to develop a user-friendly query tool to meet the needs of health data users through the application of best practices and modern technologies. The department leveraged its data warehouse architecture and set out to standardize the way data is managed and disseminated in order to ensure its confidentiality, security and integrity. The department then designed a public facing web-based interface to allow both internal and external users to easily select and drill down to the data and statistics needed.

The new web-based application, Enterprise Data Dissemination Informatics Exchange (EDDIE), was completed in December 2015. The application allows a wide array of stakeholders, including public health experts and executives, legislators, researchers, businesses, media, academia and concerned private citizens, to access over a dozen of the most popular health datasets available from the department, such as vital records, cancer incidence, behavioral health risk and numerous communicable diseases. The application is easy to navigate and includes help files linked to every selection menu and table column. The site also includes a frequently asked questions section, a short instructional video, and the ability to export data in a variety of formats.

EDDIE sets the standard and serves as a model for sharing and accessing quality health statistics in Pennsylvania. Overall, EDDIE is elevating the state's open data policy on detailed aggregate health statistics, while maintaining confidential patient information and utilizing modern technologies to meet the department's business needs.

Business Problem and Solution:

Problem

Program offices within Pennsylvania Department of Health and external public health professionals need to efficiently access and analyze health and health-related data for use in making policy and program decisions. The data and statistics must be of high quality to ensure confidence when making or influencing important decisions.

The department's approach to providing health-related data has evolved over the years. Beginning in 1990's, the department's Health Statistics area would process data requests manually. After receiving multiple requests for the same or similar data, the department began looking at opportunities for self-service by placing statistical reports and data tables online. These reports and tables were typically provided as Portable Document Files (PDF's). Over time, the department accumulated hundreds of documents and web pages and developed several different independent applications to make online dissemination of data and statistics possible. Even with these improvements, the process continued to be cumbersome.

In 2003, Pennsylvania and the state of Washington collaborated to define a menu-driven tool to allow users to search for and narrow down the data they wanted from a list of datasets. The tool, Epidemiological Query and Mapping System (EpiQMS), allowed users to build tables, maps, charts and county assessments from nine of the most commonly used public health data sources. That tool now has been in place for over 10 years and its technology is reaching end-of-life.

Solution

In 2014, the department recognized the need to take an enterprise approach to self-service data to meet the business needs of internal programs as well as external public health professionals. The solution would need to maximize the reach and utility of our rich data assets while ensuring the confidentiality and security of personal health information.

Teams within the Division of Health Informatics began planning for an enterprise-wide application. The first phase of the project focused on enhancing data within the aging EpiQMS application before bringing it into the new data warehouse architecture. The EpiQMS architecture was quickly becoming outdated and its potential failure threatened to create a huge burden on staff resources to respond to data requests.

The department also identified the need for a new menu-driven interface design to make the data selections as easy as possible for users. The initial design team consisted of database experts, application developers and statisticians/subject area experts who met weekly to assemble data dictionaries, attribute tables, wireframes, functional requirements, database design, star schemas, help files, bug lists, etc. The project utilized a standard three environment

approach – development, user acceptance testing (UAT) and production – to ensure that each new dataset or functionality was thoroughly tested. To evaluate and assess its design, the project included beta testing and feedback from actual data users in the field before it was promoted to a production environment. With the first phase completed, the department has developed the processes needed for adding new datasets and functionality that helps save resources.

The project relied on limited funds from different program areas to acquire the services of an expert in SQL Server databases, SSRS, SQL Server Analysis Services (SSAS), .NET, and Server Integration Services (SSIS). Additionally, an in-house team of approximately 10 staff members met weekly and contributed to the project as needed. The project was managed using a combination of standard techniques for high level visibility and techniques from Agile Data Warehouse methodologies for iterative requirements collection, data analysis, development and testing. An internal collaboration site was used as the hub of communication between division subject matter experts (SME), business analysts and development staff to record the versions of documentation, metadata and decisions. Development was done in an iterative fashion that allowed the SMEs to review pieces of the application in weekly meetings before promotion to a testing environment; a practice that reduced risk, ensured team cohesion and ultimately produced a high quality product that was agreed upon by both Data Warehouse and Health Statistics staff. The success of the iterative methodology has led to its use for current and projected Data Warehouse projects.

The result is a new, web-based self-service tool known as “EDDIE” or Enterprise Data Dissemination Informatics Exchange. EDDIE provides a user-friendly way to quickly create custom data tables with a variety of statistics from over a dozen of the most frequently requested datasets, such as vital records, cancer incidence, behavioral health risk and numerous communicable diseases.

The application can be accessed through a variety of web browsers using desktops computers, smart phones and other mobile devices. EDDIE is menu-driven, which means the user can follow an intuitive decision-making flow to create a custom table of statistics that include frequency counts, population (or other denominator) figures, various rates or ratios, confidence intervals, significance testing results, and demographics (e.g., county, municipality, race, sex, age group, year, health event, etc.). EDDIE provides numerous help menus, answers to frequently asked questions, an active feedback survey, a video tutorial and contact information in the event that a user has questions. To promote EDDIE, the department highlighted the application on its public website and through numerous announcements via news release, Facebook, Twitter, etc. to Department of Health and other state employees, stakeholders and internal/external data users.

Significance:

EDDIE is an online self-service tool that advances the Commonwealth's goal to cultivate a more data-driven culture across state agencies to improve and streamline government effectiveness. In addition, EDDIE meets the needs of numerous health data requesters through a single application and lessens the burden on internal staff resources in terms of maintenance and updates. Furthermore, since the datasets in EDDIE are publically accessible, it creates transparency and openness for aggregate health statistics and offers the user a variety of data formats including HTML, PDF, CSV, PIPE delimited text, Excel, and Word. Users can filter and sort the data to meet their needs and have instant access to definitions for each data item. Many of the datasets include geographical attributes that allow the end user to focus on localized statistics for various counties, regions, districts and municipalities.

Because EDDIE is an enterprise-wide application, it promotes data sharing across the entire department and helps to reduce isolated data silos and numerous applications depending on various technologies. Additional features for charts, maps, and county assessments are already planned and being developed as future enhancements.

The EDDIE application is able to reach a broad audience because it covers a wide variety of health-related topics in its data categories. Data from each of these categories are often used to satisfy legislative and media requests, epidemiological studies, emergency planning and response, health promotion and education to reduce risk, disease surveillance and burden studies, and health assessments, as well as public inquiry. Currently, EDDIE has the following data categories:

Behavioral Health includes resident adult statistical data relating to alcohol use, arthritis, asthma, cancer screening, cardiovascular diseases, diabetes, education, insurance coverage, HIV/AIDS, immunization, injury, social economic status, tobacco use, etc. These data are collected annually through the CDC sponsored Behavioral Risk Factor Surveillance System.

Demographics data are used in nearly every aspect of health data in Pennsylvania and include population estimates by age, county, ethnicity, and race.

Diseases include data on cancer incidence as well as numerous communicable and sexually transmitted diseases.

Environmental Health is a growing topic and currently includes data based on hospitalizations, lead poisoning, childhood poverty and housing by year built.

Vital Statistics include data on births, deaths, infant deaths, reported pregnancies and reproductive outcomes. These measures are particularly important for public health surveillance, epidemiological studies and conducting health assessments.

Having a robust, publicly accessible data dissemination system provides health data users with a centralized location for accessing and analyzing the data they need, while reducing the resources needed by the department to support and update the data. By streamlining these functions and datasets, the new portal not only saves time with preparing and disseminating data, but it also raises the visibility of the datasets themselves. By preparing and sharing the data for everyone to use and learn from, we increase the value of all of the data collected and provide a solid foundation for making meaningful decisions that will enrich the lives of all citizens.

The EDDIE project aligns with the Department's Strategic Plan by supporting teamwork, integrity, quality, internal and external collaboration, ease of doing business, communication, cross-functional data and reusable structures for rapid onboarding of data.

Impact and Benefit:

In order to improve and protect public health, we need to empower health professionals by expanding and enhancing their ability to use data and statistics to drive public health policy and program decisions. The EDDIE system enhances the profile of public health data by ensuring the confidentiality, security and integrity of health data while maximizing its reach and utility. It also fosters transparency in state government by allowing users access to health data and statistics in a very functional platform. The EDDIE system provides users with an intuitive way to search, manipulate, filter and sort various statistics to best meet their particular needs. The intuitive online system requires no special skills to access and analyze a wide variety of health-related statistics. As users become accustomed to EDDIE, the department expects to realize a reduction in requests for customized data and statistics.

With more knowledgeable health data consumers, we envision a more participatory and collaborative government. Through EDDIE, public health professionals as well as concerned citizens are empowered more than ever to help address community, county and state health problems. Providing department program staff and the general public with access to high-value data significantly enhances the ability to address health-related issues for the good of the public and increases the public's knowledge of the department's operations.

The data warehouse environment allowed us to expand several datasets without size limitations or decreases in query speed. Cancer incidence data was a good example of the expansion. In the previous EpiQMS system, an estimated 3.5 million new statistical records were added each year to the cancer data. With EDDIE, the number of new statistical records added each year has been expanded to 58 million for cancer. That's nearly a billion statistical records if we count historical years of cancer data currently available in EDDIE!

Data users can access statistics through EDDIE very quickly using menus, eliminating the need for them to spend hours paging through lengthy reports and numerous webpages. Based on our Google Analytics measures, the average session in EDDIE was approximately 10 minutes on a typical day. Currently we estimate 150 users access EDDIE each week and over 35 percent are new visitors. The number of visits to EDDIE has continued to grow since its launch and currently about 20-25 percent are Commonwealth employees. Although the majority of EDDIE users are from Pennsylvania (approximately 90 percent), users from just over 30 states have used EDDIE during the first quarter. In addition, users of EDDIE in Pennsylvania encompassed over 150 cities. As you can imagine, it would not be feasible for statistical staff to assist this many data users without a strong tool like EDDIE.

The EDDIE system was the product of both data and cost sharing. It embodies the spirit of the CDC's request to share funding across multiple areas for mutual benefit. This sharing reduces the burden on any single grant or program. Although the seed funding for the project was from the Cancer Registry, the pattern of sharing has been across grants from Environmental Public Health Tracking, Sexually Transmitted Diseases, Tuberculosis, HIV Partner Services and Public Health Emergency Preparedness. This cost sharing model for EDDIE and the Enterprise Data Warehouse as a whole has helped to model and support the department's move towards evidence based actions with improved health outcomes.

User feedback on EDDIE has been overwhelmingly positive and substantiates the benefits gained by the Commonwealth of Pennsylvania. Since its launch, various programs within the department (Tuberculosis and Sexually Transmitted Diseases, Infectious Disease Epidemiology, Cancer Prevention and Control, Tobacco Prevention and Control, Cardiovascular Risk Reduction, Violence and Injury Prevention, etc.) have used the application to assist with developing program interventions, identifying emerging issues in injury prevention, assessing progress toward Healthy People 2020 objectives, assessing behavior change over time and refining public health education and promotional programs. Some of the program areas also indicated that they had directed numerous external partners and data requestors (general public, media and researchers) to EDDIE.

As the demand for and usage of data continues to grow, robust data dissemination systems such as EDDIE are needed. These online systems allow us to continue to improve and streamline government effectiveness, and to evolve into more data-driven organizations that provide important, useful information and services to the public.