Distribution Agreement

Signature:

____ ______________

Edward S. K. Esser Date

Public Health Informatics: Advancing Healthcare at the Patient and Population Level

By

Edward S. K. Esser

Degree to be awarded: MPH

Executive MPH

____

Laura M.D. Gaydos, PhD
Committee Chair

____

Melissa (Moose) Alperin, EdD, MPH, MCHES
Chair, Executive MPH

Public Health Informatics: Advancing Healthcare at the Patient and Population Level

By

Edward S. K. Esser
B.S., Georgia Institute of Technology
2010

Thesis Committee Chair: Laura M. D. Gaydos, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Applied Public Health Informatics
2018

Abstract

Public Health Informatics: Advancing Healthcare at the Patient and Population Level

By Edward S. K. Esser

The increased adoption of electronic health record systems by health care providers offers an opportunity to improve public health on multiple levels. The access to quality data provided by the proposed system could improve decision making for each of these stakeholders: a patient could track and understand their latest health numbers, a physician could be alerted of a patient health condition that needs immediate attention, and health departments could design more effective interventions, thus enabling a holistic approach to public health. The deliverables within this document demonstrate how health informatics can advance public health at both the population level as well as the patient level.

The first set of deliverables address patient health. Health literacy is the ability to find, to understand, to use, and to communicate basic health information and services in order to make appropriate health decisions. Although government websites such as CDC.gov and Cancer.gov are highly reliable, they are not always the primary sources of health information for average internet users. There is a need to bridge the gap between the scientific and technical users of these federal agency websites and the average internet user. This need can be addressed through enhanced user engagement and user experiences that provide access to pertinent public health information.

The second set of deliverables address population health. The collection, analysis, and dissemination of disease surveillance data must be improved. Public health informaticians should strive to develop centrallized surveillance systems similar to a health information exchange in which surveillance data can be automatically captured and securely accessed amongst all participating stakeholders. As these surveillance systems mature, advanced predictive and prescriptive analytics will offer data scientists increased opportunities to aid in evidence-based preventive efforts. Advanced decision support from cognitive computing engines, cluster analysis, machine learning, natural language processing, and text analytics can help providers recognize diagnoses that might remain elusive otherwise, while population health management tools can highlight those most at risk of being readmitted to the hospital or developing costly diseases and infections.

Public Health Informatics: Advancing Healthcare at the Patient and Population Level

By

Edward S. K. Esser

B.S., Georgia Institute of Technology

2010

Thesis Committee Chair: Laura M. D. Gaydos, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Applied Public Health Informatics

2018

Table of Contents

Introduction

Health information technology (HIT) such as electronic health record (EHR) systems and health information exchanges (HIEs) can benefit patients, clinicians, and public health officials. These technologies can improve the quality of medical care, streamline work and decision making processes, and engage patient involvement, thus, leading to better patient and population health outcomes [1, 2].

In the United States, the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 sought to promote the adoption and meaningful use of HIT. Since it was enacted, the adoption of EHR systems that meet the technological capability, functionality, and security requirements specified by the Department of Health and Human Services has steadily increased in hospitals and office-based physician practices [3, 4].

The increased adoption of EHR systems by health care providers offers an opportunity to improve public health on multiple levels. The access to quality data provided by the proposed system could improve decision making for each of these stakeholders: a patient could track and understand their latest health numbers, a physician could be alerted of a patient health condition that needs immediate attention, and health departments could design more effective interventions. Thus enabling a holistic approach to public health [5].

The deliverables within this document demonstrate how health informatics can advance public health at both the population level as well as the patient level.

Informatics to Improve Health Literacy

Background

Health literacy is the ability to find, to understand, to use, and to communicate basic health information and services in order to make appropriate health decisions [6]. In the U.S., less than one-fifth of adults are considered to be “proficient” in health literacy [7]. Low health literacy is associated with negative health outcomes including greater rates of hospital admission and emergency room visits as well as overall poorer health. From an economic perspective, low health literacy is a major source of inefficiency in the U.S. healthcare system. It is estimated that cost of low health literacy to the U.S. economy is in the range of $106 billion to $238 billion [8]. This represents between 7% and 17% of all personal health expenditures

When designing solutions for improving health literacy, it is essential to understand the behaviors that users exhibit. According to a 2017 study [6], the average person looks for health information when they need it rather than continuously checking over time. A person seeks out health information to educate themselves when they, a family member, or friend is experiencing a change in health status. At that moment, they most likely access health information from the web browser on their mobile device (as opposed to downloading an app or waiting until they have access to a computer) and need to know relevant key words or terms (e.g., disease names or symptoms) to search. Once searched, a person needs to be able to scan and navigate through large blocks of health texts. They should be able to narrow searches so that they focus on information pertinent to the individual. Finally, a person should assess the credibility of the information. This varies depending on the level of health literacy a user exhibits. Medium and high-level health literacy users corroborate information across sources as well as determine the motivations and credentials of the content author or organization. Low-level health literacy users are less stringent in assessing the credibility of information. They may evaluate information based on if it is easy to understand, if they already agree with it (i.e., confirmation bias), or if the site is well designed.

Although government websites such as CDC.gov and Cancer.gov are highly reliable, they are not always the primary sources of health information for average internet users. According to online metrics, these sites are visited by a disproportionate amount of users with college and graduate level degrees [9, 10]. This is indicative that these sites may be too scientific-centric and, therefore, not as easy to use nor understand for the majority of the general population. There is a need to bridge the gap between the scientific and technical users of these federal agency websites and the average internet user. This need can be addressed through enhanced user engagement and user experiences that provide access to pertinent public health information.

CDC Compass

Overview

The mission of the Centers for Disease Control and Prevention (CDC) is to improve the health security of the United States and to educate the public on how to protect itself from health threats. In 2016, the CDC website had approximately 908 million page views with mobile views accounting for more than one-third of this total [11]. Traffic from mobile devices is rapidly expanding with the amount of mobile page views having increased by 50% from 2015 to 2016 [11]. As stated previously, the CDC website is currently visited by a disproportionate amount of internet users who are healthcare professionals that hold advanced college degrees. Increasing the accessibility of valuable information found on CDC.gov is imperative to improve national health literacy.

CDC Compass is an information portal that allows external users to rapidly locate and access important public health information that currently exists on CDC.gov. This public health information is critical for the development of health behaviors (health lifestyle, diet, exercise), the reduction of health risks (avoiding visits in areas with local outbreaks, improving vaccination coverage), and the communication of important public health messages (Zika outbreak, foodborne outbreaks and product recalls, etc.). CDC Compass compiles the most searched information on the CDC website and centralizes it into a single, user-friendly interface. Since more than 95% of Americans own a cell phone (more than 77% of smartphone users) [12], this portal was built with the mobile user in mind. The CDC Compass will allow the CDC to justify their mission through increased user engagement and experience through pertinent public health information. An informed population will result in improved overall health outcomes.

Features

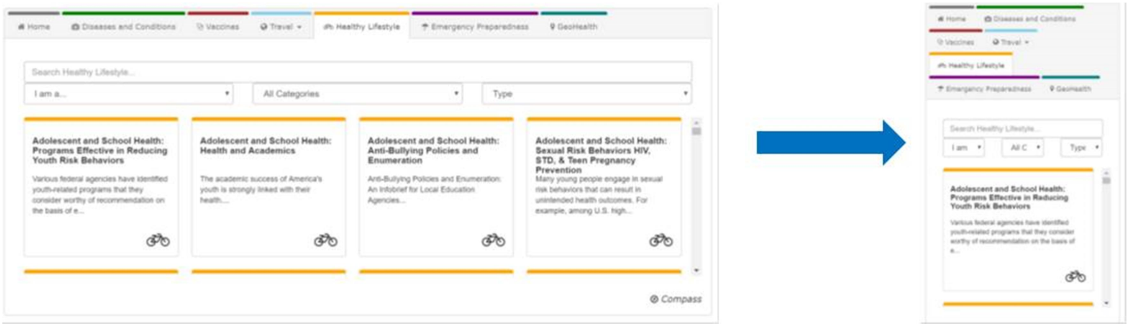
The underlying goal of CDC Compass is to enable the general public to easily find information relevant to their health and in turn, understand, use, and communicate that information.

CDC-Compass incorporates the following features:

- A mobile-friendly, responsive platform
- An optimized search engine
- A filtered and curated information display based on user characteristics, location, and needs
- An interactive travel warning map
- Integrated location services to curate the most relevant public health information (outbreaks, natural disasters, etc.) based on user location

Mobile-Friendly Platform

CDC-Compass was created with the Bootstrap framework (Fig 1). This allows for a responsive design that ensures compatibility with all screen sizes, including smartphones and tablets.



Adolescent and School Health: Programs Effective in Reducing Youth Risk Behaviors

Various federal agencies have identified youth-related programs that they consider worthy of recommendation on the basis of e...

Adolescent and School Health: Health and Academics

The academic success of America's youth is strongly linked with their health....

Adolescent and School Health: Anti-Bullying Policies and Enumeration

Anti-Bullying Policies and Enumeration: An Infobrief for Local Education Agencies...

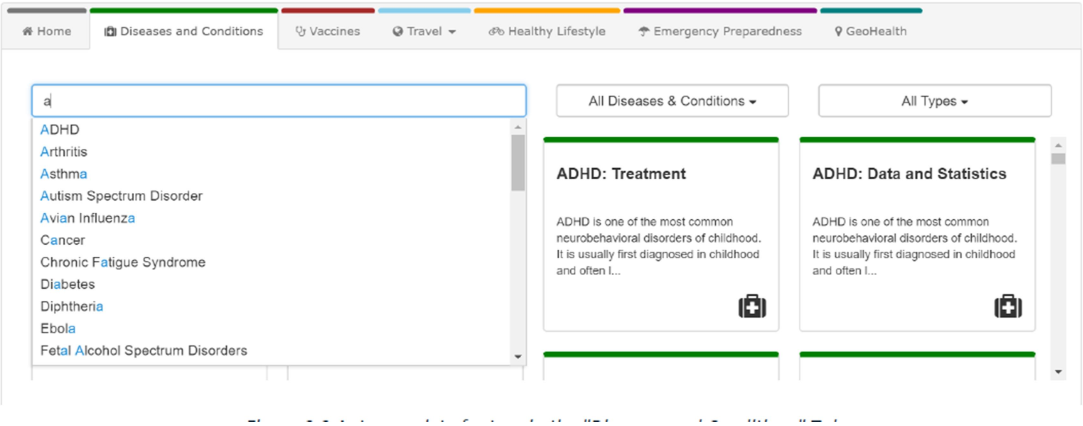
Adolescent and School Health: Sexual Risk Behaviors HIV, STD, & Teen Pregnancy Prevention

Many young people engage in sexual risk behaviors that can result in unintended health outcomes. For example, among U.S. high...

Figure 1: Bootstrap design allows for content tiles to fit any browser screen size

Optimized Search Engine

The search functionality in the CDC Compass displays tiles that link out to relevant health information (Fig 2). The portal can autocomplete searches as well as take into account the use of colloquial terms (e.g., the use of "Bird Flu" instead of "Avian Influenza").



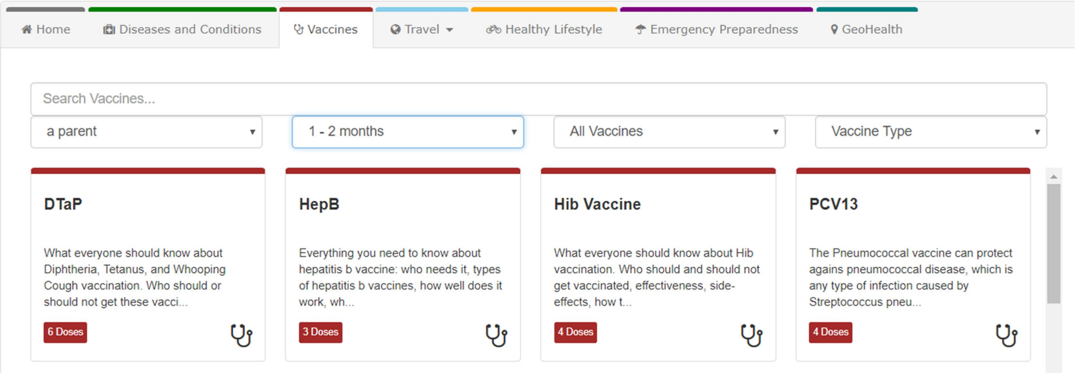
ADHD is one of the most common neurobehavioral disorders of childhood. It is usually first diagnosed in childhood and often l...

ADHD is one of the most common neurobehavioral disorders of childhood. It is usually first diagnosed in childhood and often l...

Figure 2: Portal search functionality

Filtered and Curated Information

CDC Compass allows users to quickly access information related to their own demographics through a comprehensive filter system that provides personalized information without identification (Fig 3).



DTaP

What everyone should know about Diphtheria, Tetanus, and Whooping Cough vaccination. Who should or should not get these vacci...

6 Doses

HepB

Everything you need to know about hepatitis b vaccine: who needs it, types of hepatitis b vaccines, how well does it work, wh...

3 Doses

Hib Vaccine

What everyone should know about Hib vaccination. Who should and should not get vaccinated, effectiveness, side-effects, how t...

4 Doses

PCV13

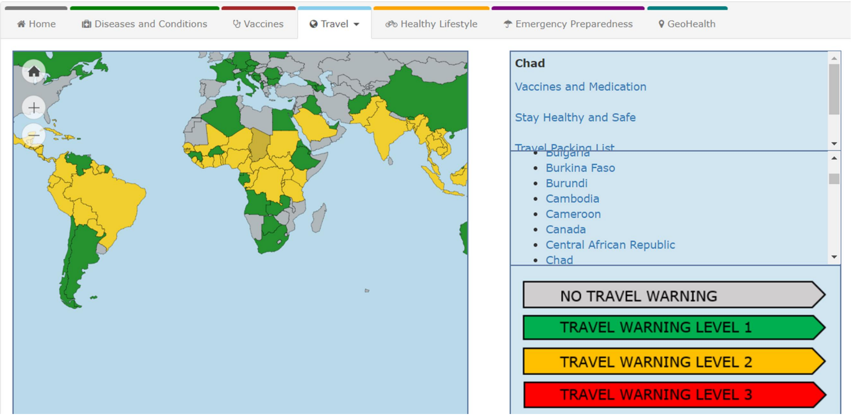
The Pneumococcal vaccine can protect agains pneumococcal disease, which is any type of infection caused by Streptococcus pneu...

4 Doses

Figure 3: Drop-down menus for user demographic information

Interactive Maps and Displays

An interactive, travel map can provide travel warnings for specific countries or regions. This map makes use of the travel warning notices color code found on the CDC website (Fig 4).



Chad

Vaccines and Medication

Stay Healthy and Safe

Travel Packing List

- Bulgaria
- Burkina Faso
- Burundi
- Cambodia
- Cameroon
- Canada
- Central African Republic
- Chad

NO TRAVEL WARNING

TRAVEL WARNING LEVEL 1

TRAVEL WARNING LEVEL 2

TRAVEL WARNING LEVEL 3

Figure 4: Travel Map

Location Services

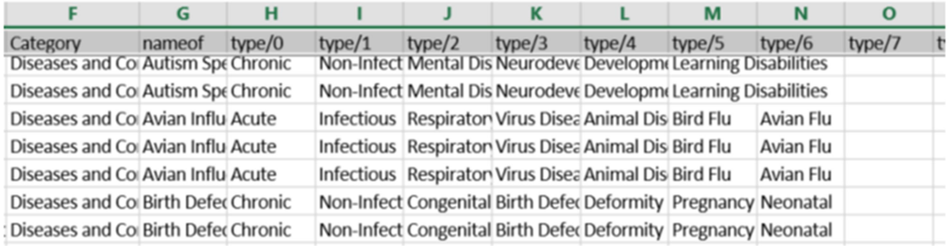
The home tab of this web portal display tiles relevant to each individual user. For example, a user in south Florida is more likely to see information related to Zika virus and hurricane preparation compared to a user in Montana. This feature utilizes truncated IP addresses and RESTful APIs that query social media sites for trending topics.

Information Architecture

Data Standardization

To enhance search functionality, content within the CDC database was standardized. A taxonomic structure was implemented to categorize public health information across the CDC enterprise. This taxonomy was based on NIH's Medical Subject Heading (MeSH) vocabulary which indexes life science journal articles for MEDLINE and PubMed as well as the clinical standards: SNOMED-CT, LOINC, and ICD-10 (Fig 5). The taxonomy arranges subject descriptors and related terms into a hierarchy. Public health content on the CDC website was then tagged with the appropriate terms (Fig 6).

Figure 5: Standards used to create metadata taxonomy



F	G	H	I	J	K	L	M	N	O
Category	nameof	type/0	type/1	type/2	type/3	type/4	type/5	type/6	type/7
Diseases and Co	Autism Spe	Chronic	Non-Infect	Mental Dis	Neurodeve	Developm	Learning Disabilities		
Diseases and Co	Autism Spe	Chronic	Non-Infect	Mental Dis	Neurodeve	Developm	Learning Disabilities		
Diseases and Co	Avian Influ	Acute	Infectious	Respirator	Virus Disea	Animal Dis	Bird Flu	Avian Flu	
Diseases and Co	Avian Influ	Acute	Infectious	Respirator	Virus Disea	Animal Dis	Bird Flu	Avian Flu	
Diseases and Co	Avian Influ	Acute	Infectious	Respirator	Virus Disea	Animal Dis	Bird Flu	Avian Flu	
Diseases and Co	Birth Defe	Chronic	Non-Infect	Congenital	Birth Defe	Deformity	Pregnancy	Neonatal	
Diseases and Co	Birth Defe	Chronic	Non-Infect	Congenital	Birth Defe	Deformity	Pregnancy	Neonatal	

Figure 6: Web content was catalogued in a spreadsheet and tagged with relevant health terms

The process of classifying public health content on the CDC website could be automated using an index API and Natural Language Processing. Metadata from URLs could be extracted and analyzed. Each object would then be assigned a unique identifier and labeled with subject headings and related terms.

Logic Architecture

On the front-end, the user will access CDC Compass from the CDC.gov homepage. They will search or filter for keywords using the portal search bar or drop-down menus. The portal is built on a RESTful API. It queries a standardized CDC database using search terms and trending health topics. Finally, it retrieves metadata content to display as tiles (Fig 7).

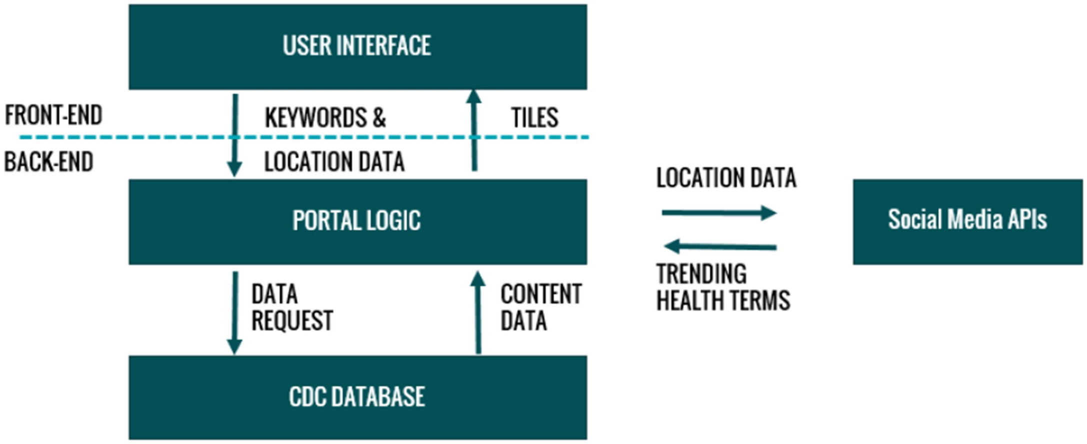


Figure 7: High level portal architecture

Cancer Care

Overview

In the year 2020, it is estimated that there will be approximately 18 million cancer survivors in the United States and that the national cost of cancer treatment will be approximately $150 billion [13]. Cancer patients and their families are faced with complex information and treatment decisions. These patients will need to perform tasks such as analyzing relative risks and benefits, calculating dosages, communicating with health providers, and interpreting test results. Cancer patients with limited health literacy are more likely to have misconceptions regarding their diagnoses and treatments and less likely to adhere to treatment plans [14]. In contrast, it has been found that when patients are actively engaged in the decision-making process regarding their health, they experience a higher quality of life compared to those patients who undertake a passive role [15]. Survey data has shown that only 50% of cancer patients are offered a part in the decision-making process[16].

The mission of the CDC's National Comprehensive Control Program (NCCCP) is to assist stakeholders at all levels in preventing or minimizing the impact of cancer in communities. One of the ways healthcare providers can provide comprehensive cancer control is to utilize existing information resources in order to support the needs of cancer patients and their families. In order to make informed health decisions, patients must be able to obtain, process, and understand cancer-related health information and services.

The National Cancer Institute (NCI) is the country's leader in cancer research. Its website, Cancer.gov, has a wealth of knowledge for nearly every facet of the disease. Both patients and physicians should be able to access this information when they view the patient's electronic medical records.

Cancer Care is a patient-friendly application built on the HL7 SMART on FHIR platform that would be jointly developed by NCCCP and NCI. This application would integrate into the EMR system of healthcare providers (Fig 8 and Fig 10) with the purpose of disseminating information regarding cancer symptoms, risk-factors, treatments, and healthy behaviors that a patient can more easily access and understand. This information would be located within the patient's healthcare portal and tailored to their diagnostic and demographic data.

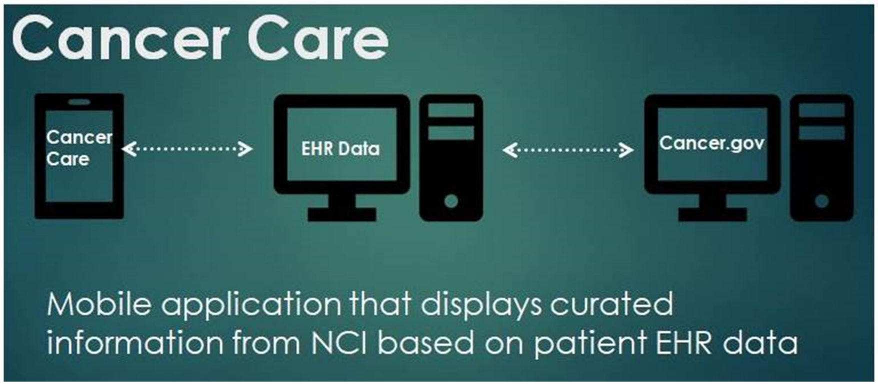


Figure 8: Cancer Care general schema

Development Plan

Cancer Care will improve the patient's understanding of treatment options by making information relating to their conditions easier to access. This process will take place on the following fronts:

- Cancer related content within the NCI website will be standardized to enhance search functionality. A taxonomic structure will be implemented to categorize public health information across the NCI enterprise. This taxonomy will be based on existing health standards (ADT, SNOMED-CT, LOINC, etc.) that are incorporated into HL7 messages. Cancer content on the NCI website will be analyzed using natural language processing and machine learning technology. Metadata for URLs on the site will be tagged with relevant keywords and standard codes.
- A SMART of FHIR HL7 application will be integrated into a healthcare provider's EHR system. This application will read the patient's HL7 health data and query the NCI database for relevant cancer information. The application will then present this relevant information to the patient or Cancer Care user.
- A user-friendly interface to display cancer related information will be developed. The application must employ best user experience practices for displaying this information. It must be presentable on devices and browsers of varying size.
- A pilot project will be arranged between NCI and the Winship Cancer Institute (WCI) at Emory University. The Winship Cancer Institute was selected because of its world renowned reputation, use of a Cerner EHR system, and proximity to the CDC.

Enterprise Architecture

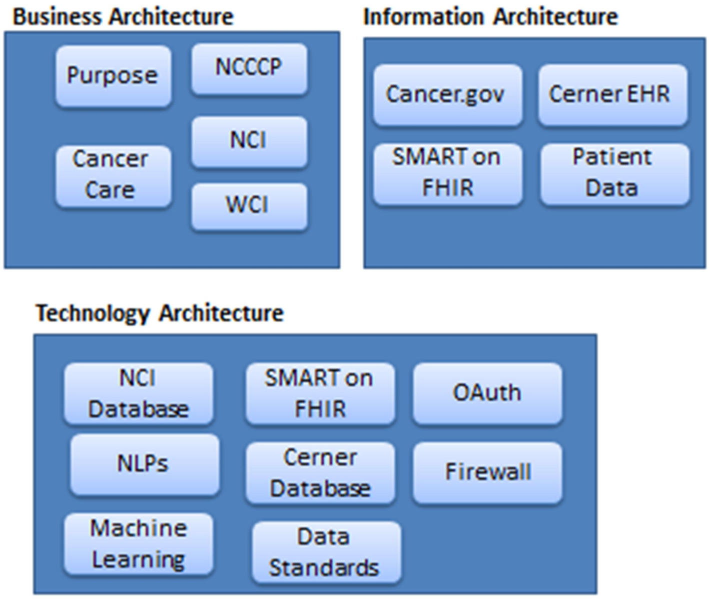


Figure 9: Cancer Care enterprise architecture schemas

Business Architecture

Cancer Care will allow patients to have a better understanding of their treatment options allowing them to have increased satisfaction and engagement with their physician. Improved health outcomes will justify the mission of the NCI and Winship Cancer Institute.

Information Architecture

Cancer Care will act as a translator between the NCI's Cancer.gov website and Winship Cancer Institute's EHR system. Patients and physicians will be able to access the information on the NCI website. Likewise, NCI will be able to see what topics relating to cancer are most relevant to treatment decisions.

Technology Architecture

NCI content will be standardized and classified using NLP and machine learning technologies. On the patient side, HL7 messages on the Cerner database will be securely sent to the Cancer Care SMART on FHIR platform. These messages will be analyzed for health standard data. Identified standards will be used to query the NCI database and content related to the standards will be sent to the Cerner Patient Portal.

Logic Architecture

Represented in the diagram below (Fig 10), information within the EHR-NCI network would flow as follows:

1. A patient is admitted into the hospital or clinic.
2. As they undergo treatment/testing, their electronic medical record data is updated.
3. Cancer Care analyzes the patient's medical record data for standard codes related to cancer.
4. Standards related to cancer will be used to query a structured National Cancer Institute database.
5. Cancer Care will then retrieve URL links from NCI's Cancer.gov website.
6. A user interface will display the URLs for the patient to review.

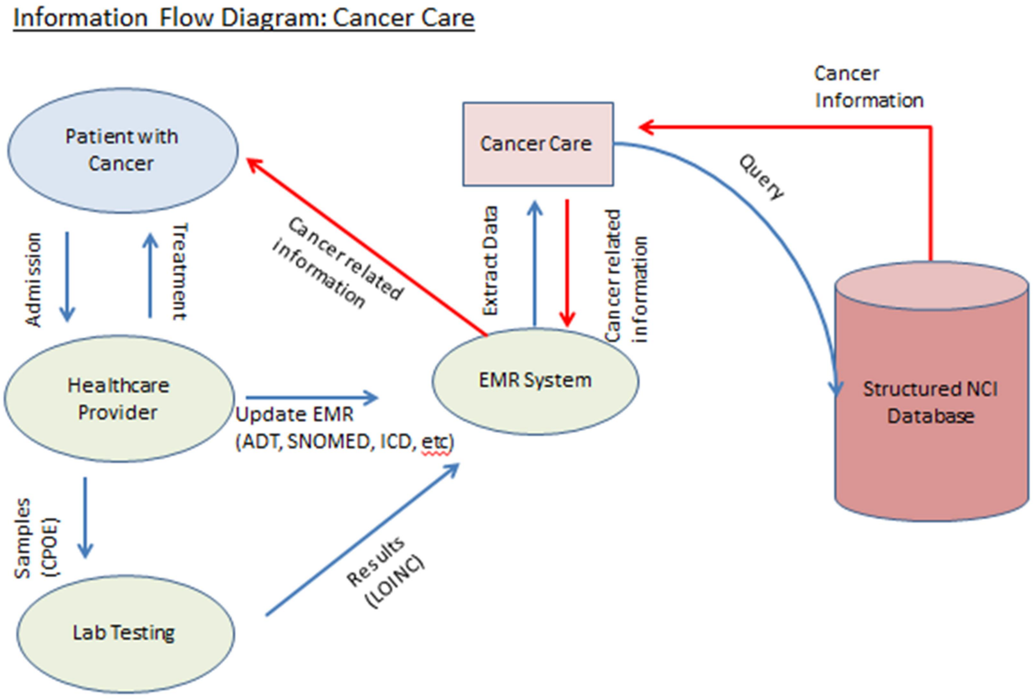


Figure 10: Cancer Care information flow

Discussion on Improving Health Literacy through Informatics

Both CDC Compass and Cancer Care are informatics solutions that could improve the user experience and health literacy; thus, potentially leading to improved health outcomes.

Effective user experience guided the CDC Compass project from its inception. An evaluation of online user behaviors on the CDC website highlighted gaps and inefficiencies in the navigation of the website (e.g., more than 3 clicks to find information, non-functional links, disparate server capabilities). CDC Compass was built and designed to optimize user experience which will enable the CDC to effectively disseminate public health information to a wider audience. One user experience characteristic the project focuses on is search functionality. The new search functionality is driven by standardized meta-data which improves user experience by considering variation in possible searches.

CDC Compass also provides the CDC with an opportunity to implement a standardized metadata taxonomy across all centers. This would increase interoperability across the enterprise, and elevate the overall visibility of the Centers. CDC Compass was designed with scalability in mind and the underlying processes involved in developing the portal could also translate to other HHS organizations such as the Federal Drug Administration, Centers for Medicare and Medicaid Services, National Institute of Heath, and National Cancer Institute (NCI).

The adoption of EMRs by HCPs has led to improved quality, safety, efficiency of treating patients and thus reducing health disparities amongst populations[17].By making health information more accessible, EMRs have also allowed HCPs to better engage with patients and their families during the treatment process [18]. However, providing access to this information is not enough. Patients also need to understand the data within their medical records and apply it to their own demographic and behavioral factors[19].

When assessing one's health results on their healthcare provider's patient portal, the portal may not relay information in a useful manner. While numbers produced from laboratory tests may be useful for a physician, a patient may not know how to contextualize them. It would be more valuable to display test results along with information on symptoms, risk-factors, or healthy behaviors that a patient can more easily understand. There is an opportunity to work with health departments, the CDC, and HCP organizations that have an EMR system in place and develop a patient portal that is more patient-centric. This would involve analyzing information within patient EMR records based on existing standards (e.g., ICD-10, HL7, LOINC, SNOMED-CT, etc.) and querying publicly available health information from state and federal public health agencies. For example, imagine a person has been diagnosed with leukemia. Their patient portal would still contain information on white blood cell levels, but it would also link out to easier to understand information from the NCI's Cancer.gov site. From this, Cancer Care was devised.

As technology advances, there will be opportunities to develop and implement strategic plans to address patient needs. Both CDC Compass and Cancer Care demonstrate how informatics solutions can bring together key partners and organizations to improve health literacy and patient centered care by utilizing existing information resources. Results for these deliverables include: promotion of healthy behaviors, access to quality health care, support for those going through various treatments.

Informatics to Improve Disease Surveillance

Background

Public health surveillance in the United States has existed since the colonial era, but national surveillance began in the year 1850 when the federal government began to collect and publish mortality statistics from the decennial census [20]. Weekly reporting of state communicable diseases began in 1893, but participation became more prevalent following the flu and polio epidemics in the 1910s. In 1952, the precursor to the Council of State and Territorial Epidemiology (CTSE) produced the first list of nationally notifiable diseases and conditions that should be reported to the Communicable Disease Center (later named the “Centers for Disease Control and Prevention”). However, it was not until 1990 when CTSE produced a uniform case criteria for diseases that should be reported to the National Notifiable Disease Surveillance System (NNDSS) [20].

EHR systems allow for the storage of patient health data. This can include: demographic information, previous diagnoses, laboratory test results, prescriptions, radiological images, vaccinations, and clinical notes. These features are promising and can improve patient safety, quality of care, operational efficiency, and cost of healthcare [21].

The collection, analysis, and dissemination of data must be improved. Public health informaticians should strive to develop surveillance systems similar to a health information exchange in which surveillance data can be automatically captured and securely accessed amongst all participating stakeholders [22, 23]. These systems would allow for better access to real-time clinical data where health, demographic, and behavioral factors can be analyzed (especially at the local level).

National Health Safety Network Modernization Initiative

Overview

Healthcare-associated infections (HAIs) are the most common complication of hospital care. Patients contract HAIs while receiving medical care for an unrelated condition [24]. In 2002, approximately 1.7 million infections and 99,000 deaths were reported in US hospitals [25]. More than half of these cases could have been reasonably prevented with current evidence-based practices like maintaining proper hand hygiene and equipment sterilization [25]. HAIs result in additional treatments and extended stays in the hospital. It is estimated that they have an economic cost ranging from $28 billion to 45 billion each year. The U.S. Department of Health and Human Services (HHS) seeks to reduce preventable HAIs in order to build a safer, more affordable healthcare system.

The surveillance of healthcare-associated infections (HAIs) is critical for assessing how closely healthcare providers are following guidelines for safe care. In order to receive compensation from the Centers for Medicare and Medicaid Services, healthcare providers are required to report data for certain infections to a centralized surveillance database known as the National Healthcare Safety Network (NHSN) [26, 27].

The reporting requirements for these infections vary based on the type of patient (e.g., adults and children) as well as the type of medical facility (e.g., surgical ward, ICU, psychiatric ward, etc.) in which care is administered. NHSN is an internet application that is used to track HAIs in about 20,000 medical facilities across the nation [26, 27] . Data can be manually entered into the platform or transmitted directly from EHR systems [26]. Data from NHSN is stored in SQL database behind a secure data network.

The required reporting events do not capture all HAIs. Current NHSN reporting processes are limited and do not meet this predictive need [28, 29] for prevention and improvement of patient safety. Antibiotic resistance is a major public health problem [30]. According to the CDC, approximately 23,000 people die from multi-drug resistant organisms (MDROs) [30], with the majority of these deaths occurring in clinical settings. For this reason, it is crucial to identify more strains of antibiotic resistant bacteria and know the best methods for preventing and treating infections.

As new MDRO strains emerge and technology advances, the NHSN must be able to capture, integrate, and analyze new sources of data beyond what is currently required in order to improve MDRO surveillance. To achieve this, it is recommended that NHSN should move towards a more scalable and flexible data architecture. This data architecture would be a hybrid system incorporating both the existing relational database systems (RDBS) as well as the Hadoop framework. This platform would be able to analyze both the structured data already being collected as well as make new insights into MDRO surveillance from processes such as genomic sequencing and predictive machine learning [28, 29, 31].

Data Needs, Sources, and Uses

Currently, NHSN supports the use of HL7's Clinical Document Architecture (CDA). Data implementation tools extract HAI data from the healthcare provider's electronic medical record system, laboratory information, and Admission, Discharge, and Transfer data. This information results in summary reports that consist of the number of infections compared to the number of procedures taking place or patients being treated [26, 27].

Moving towards a Hadoop framework would allow for more data within medical records such as patient prescriptions or risk-factors to be collected and analyzed [29]. This information would be extracted using a SMART on FHIR application. Patients who contract HAIs may have underlying risk factors that are making them more susceptible. Machine learning could analyze these factors and notify physicians if the patient is at risk.

It is also an opportunity to expand the number of strains reported. Sample isolates collected from patients and medical facilities could undergo whole-genome sequencing in order to identify lower-frequency mutations in viral genomes that may not have been detected using traditional sequencing methods [7]. A databank could map these mutations and compare them across the country. It could be linked to other surveillance systems like the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS) which works closely with the FDA. Again, machine learning could be implemented to detect possible mutations and recommend the optimal antibiotic or treatment options to prevent the HAI.

This data would allow for improved:

- Detection of emerging trends of drug resistance
- Identification of risk factors leading to infections
- Improvement of interventions for patient safety issues through real-time analytics
- Education of patients and physicians on how to prevent MDRO infections

Enterprise Architecture

NHSN will be hosted on Amazon Web Services (Fig 11). It will utilize AWS: Elastic Map Reduce to store the data collected and will be the location for analysis. Security for these cloud instances will be provided by SOPHOS. Automatic server configuration will be set up using Puppet. Errors and failures in the system will be continuously monitored via Jenkins. Within the cloud environment, there will be a several platforms and applications that will allow NHSN to query, analyze, and report HAI events.

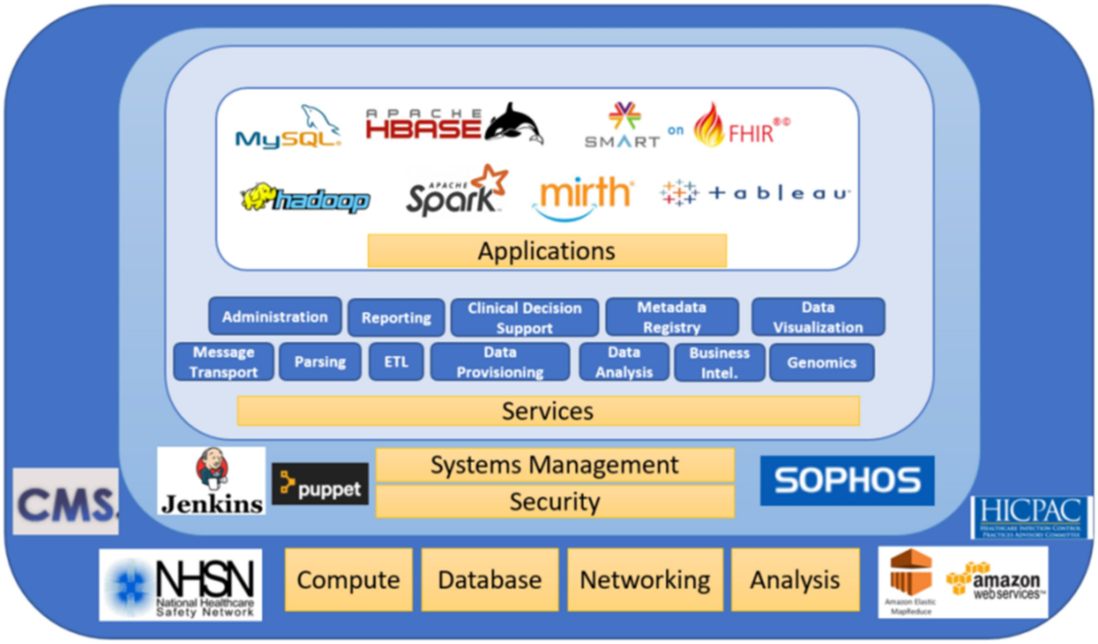


Figure 11: NHSN Modernized Enterprise Architecture

Information Flow

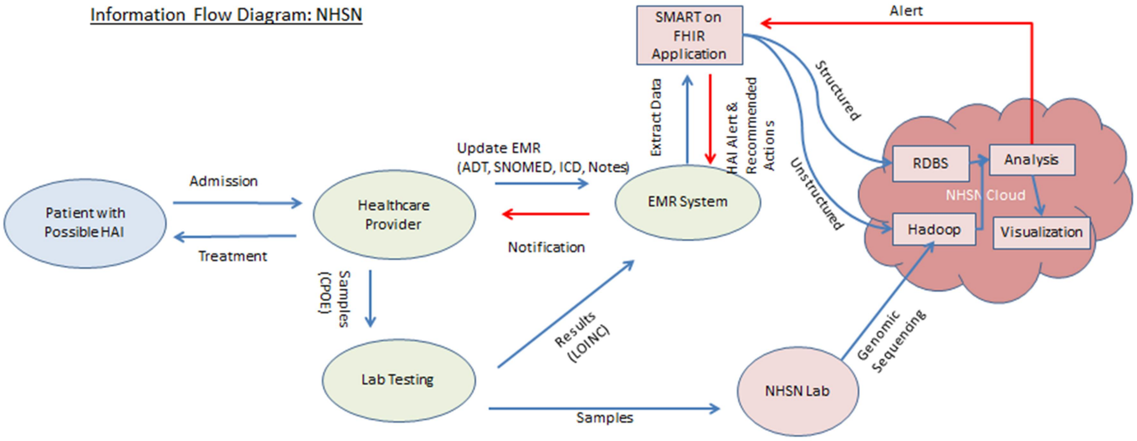


Figure 12: NHSN Information Flow Diagram

Represented in Figure 12, information within the NHSN surveillance system would flow as follows:

1. A patient (already in the hospital for at least 3 nights) would be suspected of having an HAI. People admitted to the hospital with an infection are not considered an HAI event.
2. The physician would make note of this in the patient's health records, order diagnostic tests (CPOE) from the hospital laboratory, and send patient to the ICU.
3. Bacterial organism is cultured and test results (LOINC) are updated in the EMR system.
4. If the patient has an HAI, a SMART on FHIR application would collect the information available in the patient's EMR records as well as a facility ID. This data would be categorized as structured and unstructured and sent to either the RDBS or Hadoop database within an AWS cloud storage space.
5. In addition, genomic sequencing of patient samples by a health department or associated NHSN lab would also take place. This data would be sent to Hadoop.
6. Analysis and visualization applications would be used to interpret data for public health purposes.
7. If a patient exhibits an increased likelihood for developing an HAI, an alert could be sent to the SMART on FHIR application that notifies the health care provider of the best course of treatment.
8. Health Facility reports are generated and sent to CMS.

Ithaca 2.0 Chronic Disease Surveillance System

Overview

In the United States, there is a growing population needing access to health care. An estimated 40 percent of the population has at least one chronic disease (e.g., heart disease, dementia, diabetes, cancer), which is associated with 70 to 80 percent of the health care costs in the country [32].

Of the 56.4 million global deaths in 2017, 40 million, or 70 percent, were due to chronic conditions, or non-communicable diseases (NCDs). The four primary contributors to NCDs include cardiovascular diseases, cancers, diabetes, and chronic lung diseases, which comprise 81 percent of all chronic disease deaths [33]. Generally, chronic diseases tend to be of long duration and are the result of a combination of genetic, physiological, environmental, and behaviors factors [34]. This percentage will likely continue to increase and is coupled with challenges associated with limited resources in health care and the need to better manage the health of the country's population [32].

Currently, there is no national system that can effectively provide support for federal and state partners to collect, monitor, and track patients with severe chronic conditions. There is a need for all state and federal stakeholders to amalgamate data in a centralized informatics system. By and large, these conditions (e.g., Alzheimer's disease and related dementias, diabetes, cardiovascular diseases, and chronic pulmonary diseases) are considered non-reportable; providers are not under mandate to report events to the state health department, nor are state health departments under mandate to report to the federal level.

Ithaca 2.0 is a modern surveillance platform developed to better understand disease progression through collection and analysis of chronic cases, while also providing interoperability with leading systems (e.g., CA DPH, NYSDoH, Texas DSHS). With the increasing size of the aging population, a diagnosis of chronic illness will have a greater impact on the cost of health care services if prevention efforts are not expanded. By creating an integrated national surveillance system for chronic conditions, we can collect data from all stakeholders in a timely manner, track prevalence and incidence, identify common risk factors, recommend targeted prevention efforts, improve quality of care, decrease disparities in care, identify research gaps, monitor population trends, evaluate programs, and lower the overall health care burden.

Features

Ithaca 2.0 is a national system that will provide support for federal and state partners to collect, monitor, and track patients with chronic diseases.

The objectives of this system are the following:

- Integrate clinical, epidemiological, and genomics data to support surveillance programs and translational research, transforming discoveries in genetics into practical health care application
- Increase awareness about preventing and treating chronic conditions
- Provide empirical evidence to expand successful health care programs and interventions
- Develop innovative tools, visualizations, and reports to improve access to data and share findings (Fig 13)

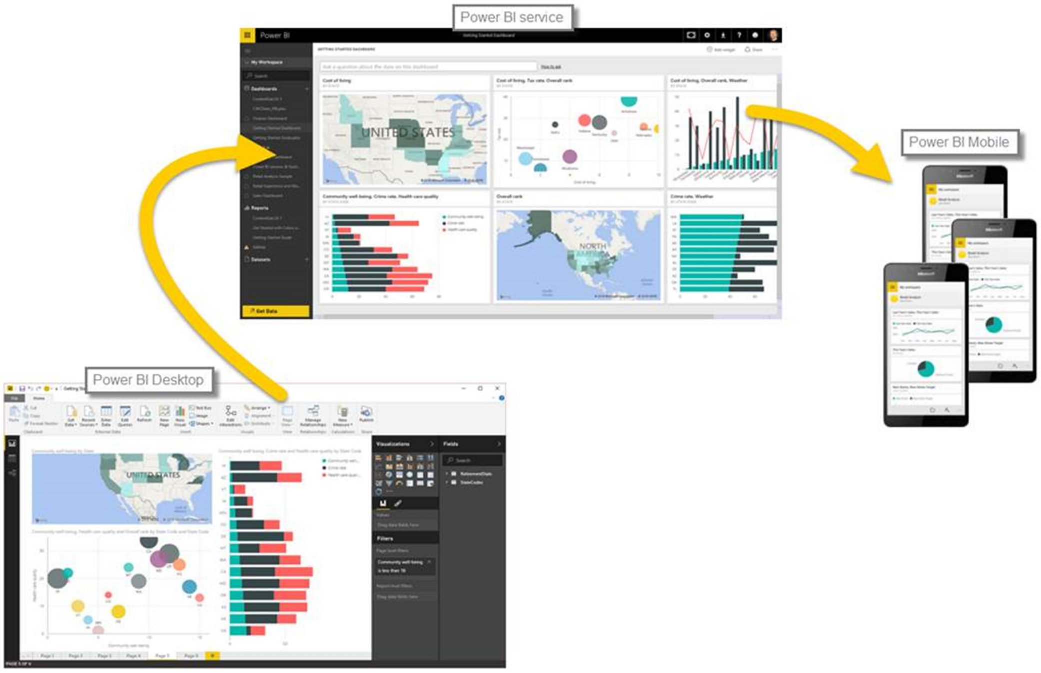


Figure 13: Ithaca 2.0 User Interface Dashboard

Enterprise Architecture

Data Flow

Ithaca 2.0 is essentially a vast, centralized clinical data warehouse – co-owned by the federal and state governments. There are several elements of the project:

- Data on chronic diseases will be stored within the clinical data warehouse.
- Data will come from state and federal agencies.
- Annually, new data will be added to flow directly into the clinical data warehouse, based on evaluation of justifying business cases by a governance board.
- Users of the clinical data warehouse will also be able to import their own data files if the data does not automatically flow into the warehouse. If their data includes a

common, unique identifier within the warehouse, they will be able to join their tables for analysis.

- An off-the-shelf data visualization tool will be integrated into the warehouse to allow for easy data analysis and visualization.

Data Architecture

Ithaca 2.0 will consolidate patient data from health information exchanges and registries throughout the country. A cloud computing infrastructure (Fig 14) will allow for enhanced efficiency, scalability, and reliability that can provide needed functionality and services for Ithaca 2.0. State and federal health organizations will be able to integrate health data and employ analysis tools within the platform.

Ithaca 2.0 will be deployed on the MS Azure Government cloud service. Health data will upload into a data repository (MS Azure Blob and Table Storage). Required data from the source systems will be extracted for processing. Data values will be converted into standardized forms/structures and be loaded into the data warehouse where analysis will take place using Azure Analysis Services.

Azure Virtual Machines provide computing resources based on workload demand. Ithaca 2.0 employs Ls-series virtual machines, which are optimized for storage and ideal for SQL and NoSQL databases.

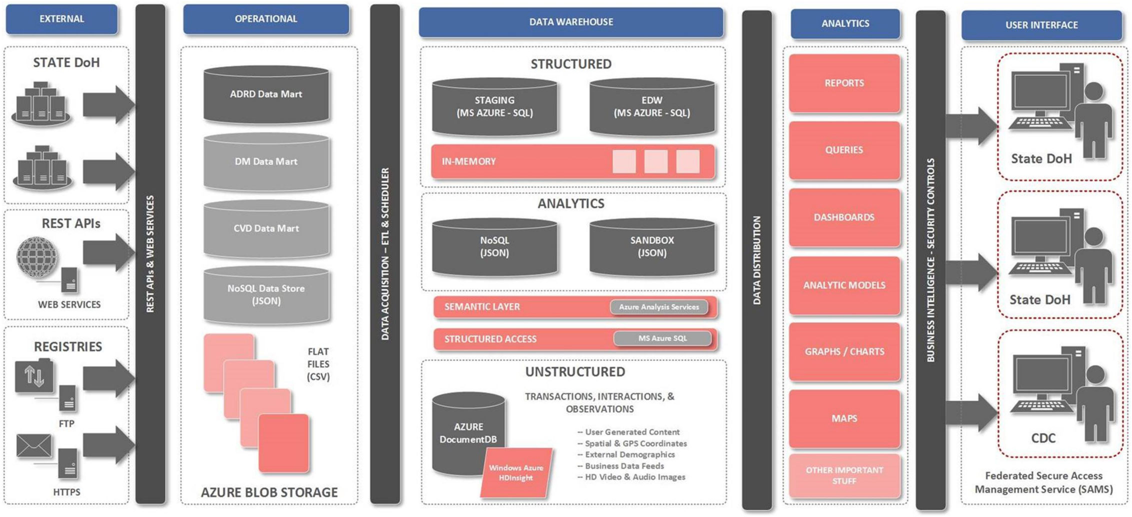


Figure 14: Ithaca 2.0 Data Architecture

Discussion on Informatics Improving Disease Surveillance

Detection, screening, and treatment of acute and chronic diseases, as well as palliative care, are key elements of the global imperative to respond urgently. Improved health care, early detection, and timely treatment are proven, effective approaches in reducing the impact of disease. However, appropriate care for people with acute and chronic conditions is lacking in lower-resource settings, and access to essential technologies and medicines is limited, particularly in low- and middle-income communities. Many health care interventions are cost-effective, especially compared to costly procedures that may be necessary when detection and treatment are late and the patient reaches advanced stages of the disease. Health systems need to be further strengthened to deliver effective, realistic, and affordable interventions and services for people with acute and chronic conditions.

The confluence of scientific breakthroughs in next-generation sequencing, cloud computing, and advanced analytics gives rise to new surveillance possibilities. Health informatics solutions will catalyze change within the context of how acute and chronic illnesses are addressed in the United States – from the fragmented, disparate approaches focused on local monitoring of diseases to an integrated, standardized central model focused. These surveillance solutions will lead to the development of enhanced diagnostics, early-intervention, and disease prevention strategies. In addition, these new technologies will enable data scientists to traverse disease domains and synthesize disparate data sets (e.g., genetic, ecologic, environmental, epidemiological, clinical, laboratory, public health, geo-spatial) to promote analytical insights and health advocacy.

The modernized NHSN and Ithaca 2.0 surveillance systems are platforms upon which to manage and monitor data streams, apply distributed and parallelized data processing across "big data" sets (e.g., whole genome sequences), and introduce business intelligence and data visualization techniques. Federal and state bioinformatics partners will be able to securely access a wide range of disease data, risk factor indicators, and policy measures to describe the burden of chronic disease as well as common risk factors, identify research gaps, monitor population trends, and evaluate programs.

As these surveillance systems mature, advanced predictive and prescriptive analytics will offer data scientists increased opportunities to aid in evidence-based preventive efforts (e.g., early detection, disease prevention and intervention, and capacity planning). Advanced decision support from cognitive computing engines, cluster analysis, machine learning, natural language processing, and text analytics can help providers recognize diagnoses that might remain elusive otherwise, while population health management tools can highlight those most at risk of being readmitted to the hospital or developing costly diseases and infections.

Conclusion

In today's digital age, an individual's data can be stored for virtually anything, and, more importantly, communicated to almost anyone. Advancements in communications technology have allowed for a new level of interconnectedness not seen by previous generations. This interconnectedness has brought about a global society that is able to easily exchange different world views, cultural aspects, commercial products, and scientific information.

The proliferation of mobile devices and wearable technologies also brings about the opportunity to increase access to healthcare and improve decision making at the individual and population levels [35]. Public health informaticians have the ability to collect data and information from different communities or populations in ways that were previously not possible. Health data analytics platform built on frameworks like Hadoop/MapReduce allow for the ingestion and integration of data from different sources, structures, and formats [36]. By capturing complete datasets, a more representative picture of a problem may be examined.

Public health informaticians can analyze this information, identify the needs of greatest importance, and help construct an intervention plan to address the specific needs and incentives of these communities. The work of an informatician never stops. There will always be new stakeholders to assist, new technologies to implement, and new problems to solve.

References

1. Buntin, M.B., et al., *The benefits of health information technology: a review of the recent literature shows predominantly positive results.* Health Aff (Millwood), 2011. **30**(3): p. 464-71.
2. Jamoom, E.W., et al., *Physician Opinions about EHR Use by EHR Experience and by Whether the Practice had optimized its EHR Use.* J Health Med Inform, 2016. **7**(4): p. 1000240.
3. Charles, D., M. Gabriel, and T. Searcy, *Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014*, in *ONC Data Brief, no. 23*, O.o.t.N.C.F.H.I. Technology, Editor. 2015: Washington DC.
4. Jamoom, E.W., N. Yang, and E. Hing, *Adoption of Certified Electronic Health Record Systems and Electronic Information Sharing in Physician Offices: United States, 2013 and 2014.* NCHS Data Brief, 2016(236): p. 1-8.
5. Levy, B.S., *Toward a holistic approach to public health surveillance.* Am J Public Health, 1996. **86**(5): p. 624-5.
6. Champlin, S., et al., *Toward a Better Understanding of Patient Health Literacy: A Focus on the Skills Patients Need to Find Health Information.* Qual Health Res, 2017. **27**(8): p. 1160-1176.
7. Davis, T.C., et al., *Health literacy and cancer communication.* CA Cancer J Clin, 2002. **52**(3): p. 134-49.
8. Berkman, N.D., et al., *Health literacy interventions and outcomes: an updated systematic review.* Evid Rep Technol Assess (Full Rep), 2011(199): p. 1-941.
9. *CDC.gov Traffic Statistics*. 2018 4/8/2018]; Available from: https://www.alexa.com/siteinfo/cdc.gov.
10. *Cancer.gov Traffic Statistics*. 2018 4/8/2018]; Available from: https://www.alexa.com/siteinfo/cancer.gov.
11. *CDC Digital Media Metrics*. 2018 4/8/2018]; Available from: https://www.cdc.gov/metrics/index.html.
12. *Mobile Fact Sheet*. 2018 4/8/2018]; Available from: http://www.pewinternet.org/fact-sheet/mobile/.
13. Mariotto, A.B., et al., *Projections of the cost of cancer care in the United States: 2010-2020.* J Natl Cancer Inst, 2011. **103**(2): p. 117-28.
14. Koay, K., P. Schofield, and M. Jefford, *Importance of health literacy in oncology.* Asia Pac J Clin Oncol, 2012. **8**(1): p. 14-23.
15. Kiesler, D.J. and S.M. Auerbach, *Optimal matches of patient preferences for information, decision-making and interpersonal behavior: evidence, models and interventions.* Patient Educ Couns, 2006. **61**(3): p. 319-41.
16. Stacey, D., L. Paquet, and R. Samant, *Exploring cancer treatment decision-making by patients: a descriptive study.* Curr Oncol, 2010. **17**(4): p. 85-93.
17. Institute of Medicine Roundtable on Health, L., in *Innovations in Health Literacy Research: Workshop Summary*. 2011, National Academies Press (US)

Copyright 2011 by the National Academy of Sciences. All rights reserved.: Washington (DC).

18. Institute of Medicine Committee on Quality of Health Care in, A., in *Crossing the Quality Chasm: A New Health System for the 21st Century*. 2001, National Academies Press (US)

Copyright 2001 by the National Academy of Sciences. All rights reserved.: Washington (DC).

19. McKenna, V.B., J. Sixsmith, and M.M. Barry, *The relevance of context in understanding health literacy skills: Findings from a qualitative study.* Health Expect, 2017.

20. Chorba, T.L., et al., *Mandatory reporting of infectious diseases by clinicians.* MMWR Recomm Rep, 1990. **39**(Rr-9): p. 1-17.
21. Ajami, S. and R. Arab-Chadegani, *Barriers to implement Electronic Health Records (EHRs).* Mater Sociomed, 2013. **25**(3): p. 213-5.
22. Thacker, S.B., J.R. Qualters, and L.M. Lee, *Public health surveillance in the United States: evolution and challenges.* MMWR Suppl, 2012. **61**(3): p. 3-9.
23. Leider, J.P., et al., *Data, Staff, and Money: Leadership Reflections on the Future of Public Health Informatics.* J Public Health Manag Pract, 2017. **23**(3): p. 302-310.
24. Whittington, M.D., et al., *Value of Public Health Funding in Preventing Hospital Bloodstream Infections in the United States.* Am J Public Health, 2017: p. e1-e6.
25. Umscheid, C.A., et al., *Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs.* Infect Control Hosp Epidemiol, 2011. **32**(2): p. 101-14.
26. CMS. *Healthcare-Associated Infections*. [cited 2017; Available from: https://www.medicare.gov/hospitalcompare/Data/Healthcare-Associated-Infections.html.
27. CDC. *National Healthcare Safety Network*. 2015 15 Oct 2015 [cited 2017; Available from: https://www.medicare.gov/hospitalcompare/Data/Healthcare-Associated-Infections.html.
28. Wright, M.O., K. Allen-Bridson, and J.N. Hebden, *Assessment of the accuracy and consistency in the application of standardized surveillance definitions: A summary of the American Journal of Infection Control and National Healthcare Safety Network case studies, 2010-2016.* Am J Infect Control, 2017. **45**(6): p. 607-611.
29. Freire, S.M., et al., *Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data.* PLoS One, 2016. **11**(3): p. e0150069.
30. Pray, L. *Antibiotic Resistance, Mutation Rates and MRSA*. Nature Education 2008 [cited 1 1]; 30]. Available from: https://www.nature.com/scitable/topicpage/antibiotic-resistance-mutation-rates-and-mrsa-28360.
31. Chung, P.Y., *Current technology in the discovery and development of novel antibacterials.* Curr Drug Targets, 2017.
32. Riippa, I., M. Linna, and I. Ronkko, *The effect of a patient portal with electronic messaging on patient activation among chronically ill patients: controlled before-and-after study.* J Med Internet Res, 2014. **16**(11): p. e257.
33. Piroth, L., et al., *Epidemiology, diagnosis and treatment of chronic hepatitis B in HIV-infected patients (EPIB 2005 STUDY).* AIDS, 2007. **21**(10): p. 1323-31.
34. Denoeud, L., et al., *Predicting pneumonia and influenza mortality from morbidity data.* PLoS ONE, 2007. **2**(5): p. e464.
35. Abbasgholizadeh Rahimi, S., et al., *Are mobile health applications useful for supporting shared decision making in diagnostic and treatment decisions?* Global Health Action, 2017. **10**(sup3): p. 1332259.
36. Cha, S., et al., *Towards a 'Big' Health Data Analytics Platform.* 2015 Ieee First International Conference on Big Data Computing Service and Applications (Bigdataservice 2015), 2015: p. 233-241.