As we all know too well, the world has entered an age of superabundant information – with devices spewing out masses of text and data ranging from consumer shopping patterns to life-saving laboratory results, weather and flight patterns and more mundane data like Facebook updates about friends and acquaintances. As the amount of information generated globally increases exponentially, we become ever more dependent on computers and information systems to make sense of it. In no area of human endeavor is this sense-making more vital than in the area of human health. A key challenge related to sense-making in human health has to do with conquering the physical and intellectual divide between clinical data and information (related to a single patient – intervention) and public health data and information (related to groups of patients, communities and populations) in support of prevention. This latter area of intellectual endeavor I have named “prevention” informatics. In this paper I will outline a number of opportunities for information scientists to contribute fresh approaches to bridging the data and disciplinary gaps that impede progress in prevention informatics today.

Why is prevention important? Over the past 100 years a health system has developed in the United States that dramatically reduces death rates –
FIGURE 1. Crude death rate for infectious diseases—United States, 1900-1996
[Adapted by Rear Admiral Dr. Patrick O’Carroll, Regional Health Administrator, U.S. Public Health Service Region X]

<table>
<thead>
<tr>
<th>Year</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900</td>
<td>1000</td>
</tr>
<tr>
<td>1920</td>
<td>800</td>
</tr>
<tr>
<td>1940</td>
<td>600</td>
</tr>
<tr>
<td>1960</td>
<td>400</td>
</tr>
<tr>
<td>1980</td>
<td>200</td>
</tr>
<tr>
<td>2000</td>
<td>100</td>
</tr>
</tbody>
</table>

Americans do not live longer, healthier lives. And Americans’ cancer survival rates are not markedly better than those of other developed countries [1]. Consensus opinion is that the United States lags in basic preventive care.

The economics of intervention versus prevention information systems are similar. There is a great deal of investment in electronic medical records systems, which track the interventions that the patient has experienced, but limited investments in the informatics of prevention. This limited investment continues in spite of ever-increasing threats of infectious disease and environmental health issues such as water and food safety resulting from the speed at which infectious agents can now disseminate around the world.

The unprecedented volume and speed of human mobility are perhaps the most conspicuous manifestations of the present era of globalization. This has been called the “Global Express” – “the system that connects us across oceans, continents, national boundaries, cultures, languages, groups, ethnicity and trade systems” [2, p. xvii]. A century and a half ago, it took about 365 days to circumnavigate the globe by ship; today, with air travel, it takes less than 36 hours. Thus the incubation period of many infectious diseases is now longer than the time it takes the infected to travel from one

How did medical care or clinical interventions do in comparison during the 20th century? Not so well! The greatest health gains for populations have derived from initiatives that had little to do with the treatment of illness (Table 1).

In addition to the vital goals of saving lives and preventing illness and injury, what are the economics of prevention versus intervention? The United States spends much more on health care than countries with similar kinds of economies, and although Americans have excellent access to high-tech diagnostic tools and to surgical procedures like angioplasties,
location to another. In the past, infectious disease outbreaks were readily detected on ships as they pulled into port, and the ships were quarantined until the diseases had burned themselves out. Now, should a local outbreak spread silently and globally via an infected traveler or tourist, cases will likely start emerging only days or weeks later in clinics and communities worldwide. The surveillance of mobile populations is only one of several types of surveillance urgently in need of development or improvement. In addition, as the transcontinental movement of food and commodities has increased, so, too, has the need for improved trade- and food-related surveillance [3]. Strategic information systems that can, in real time, assemble disparate information and data regarding trade and human movements (for example, flight records) along with weather patterns and on-the-ground surveillance for disease outbreaks are urgently needed.

Definitions: Intervention Informatics and Prevention Informatics

For most of the last 30 years the development of medical information systems has focused primarily on the electronic medical record and information systems that support clinical care such as pharmacy, laboratory and radiology information systems – that is, systems that are intervention oriented. The concept of “intervention informatics” has a particular focus on the individual, the patient with injury, disease or abnormal condition. The electronic medical record is used to track diagnoses, actions, procedures, therapies, diagnostic tests and so forth, based on episodes of care. The electronic medical record tends to be reactive – recording actions taken after the health problem occurs. It also typically lacks context related to the patient’s environment, that is, home/community (urban/rural/inner city), family members/relationships, travel or hobbies and related activities, all of which can have a direct bearing on the patient’s health.

In contrast, the concept of “prevention informatics” offers a view of the individual in context, including family, relationships and community, in support of promoting the health and well-being of the individual and populations. Aggregate information extracted from medical records can, of course, be useful for community disease prevention and response, but prevention informatics supports a view of health information systems beyond the electronic medical record, including hospital information systems that prevent medical errors and hospital-acquired infections, tracking and rapid communication approaches to respond to food contamination threats, water quality issues and the like, as well as rapid detection and response to global disease and health threats. A hallmark of prevention informatics is that it is proactive. It is highly interactive, data-intensive and data-driven, linking disparate data, resources, tools and technologies in real time.

Neither the concept of intervention informatics nor prevention informatics is in common use in the informatics community. However, they provide a useful paradigm for thinking about how one might re-design and enhance existing information systems and, indeed, create entirely new information systems and resources, to drive prevention at the individual, community and population level.

The following examples of potential prevention informatics strategies are drawn from a number of domains: classification, knowledge management and disease outbreak prevention systems.

Classification

What do these three activities have in common?

- Listing causes of death (in a 17th-century mortality table) that include “fainted in a bath,” “frighted” and “itch”
- Assigning subject headings to books in a library
- Separating machine-washable clothes from hand-washables?

All, of course, are examples of classification – upon which information systems of all types are built [4]. William Farr, in 1837 said

The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases… The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay. (Quoted in [5].)
Criticism of disease classification systems sounds eerily familiar in this quote from Dr. Jim Cimino, an expert on medical classification, 160 years later: inconsistency, lack of concept permanence, disregard for context, slow adaptation to new/emerging disease terminology – and so on [6]. To this list should be added “prevention-oriented terminology.” There are entire vocabularies devoted to

- Diseases (International Classification of Diseases - http://apps.who.int/classifications/apps/icd/icd10online/)
- Laboratory procedures (LOINC – Logical Observations, Identifiers, Names and Codes http://loinc.org/)
- Drugs (http://www.micromedex.com/index.html).

However, prevention-oriented terminology is not addressed in any systematic way. The Unified Medical Language System (UMLS) Terminology Services (www.nlm.nih.gov/research/umls/), a powerful set of vocabulary resources of the National Library of Medicine, provides, through its metathesaurus browser, a view of prevention-related terminology across hundreds of vocabularies with approximate 1100 concepts related to prevention (see Figure 2). Nonetheless, a systematic, controlled vocabulary of prevention-oriented concepts and terminology could be a useful component of clinical and public health information systems to drive the generation of alerts and reminders and to compile critical data from disparate information systems and resources for reducing disease and improving health outcomes.

Knowledge Management

The challenge: Neither the creation nor the distribution of information resources (data of all types, guidelines, research findings, maps, policies, laws, evaluation metrics, teaching materials, etc.) upon which public health practitioners depend is managed or presented in any systematic or comprehensive way at the present time. The key role of public health professionals is prevention. In fact, prevention is in the official name of the Centers for Disease Control (and Prevention). The need for rapid access to information to support critical decisions in public health cannot be disputed; however, information systems of public health professionals lag far behind those of their clinical counterparts and tend to be siloed, single-purpose systems. Examples include immunization reporting systems, which are typically not tied to medical record systems; disease reporting systems (individual diseases disconnected from each other and from clinical information systems); and stand-alone environmental health and data. A knowledge management system, developed at the University of Washington, Center for Public Health Informatics, is aimed at providing a role-driven knowledge management environment for public health officials (www.myph.org). The goal of the research is to identify approaches to removing uncertainty in public health decision making by promoting the collective sharing of information with a knowledge repository, codifying knowledge assets so they can be easily found and providing an interface which bridges information and data silos [7]. Further research in the knowledge management challenges of public health professionals is urgently needed.
Improved Approaches to Data Collection, Mining and Visualization

The classic problem is too much data and not enough information upon which to make decisions. One of the most important developments in prevention is the ability to map and visualize data across districts, nations, regions and the world, and across time – a great step forward from the groundbreaking work of Dr. John Snow in identifying the sources of the London cholera epidemic in 1854 [8]. GIS mapping and visualization, coupled with the explosive use of the Internet and mobile phones for data collection and communications, provide a rich environment for developing strategies for building information systems to anticipate and prevent global health threats. Such systems can bring together in new ways a variety of types of data not only to detect health challenges, but also to prevent their occurrence. Such data include satellite data, airline data, non-prescription drug purchases, news media, published reports from local newspapers and Internet activity, including Google concept searches and citizen-contributed information and data.

In many parts of the world, the lack of reliable infrastructure (including computing and communications) and the lack of skilled workers have made accurate and reliable data collection difficult. However, the exponential growth of cellphone infrastructure and availability is rapidly transforming data collection and management in remote areas throughout the world. A variety of types of data – from text to photos, location, blood samples, audio, barcode scans and video can rapidly be collected as well as transmitted to public health settings across the country or across the globe.

Open Data Kit (ODK) is a suite of open source tools developed by computer scientists and engineers at the University of Washington and around the world. These tools are proving to be faster and more accurate than paper-based forms for data collection and analysis and less expensive than alternative computing technologies. By using existing cellular networks, ODK’s developers are freeing the users from the traditional computer networking and related infrastructure constraints in developing countries and beyond. Features including GPS, video and photos provide a contextually richer set of data than is possible using typical paper forms, and the information can be compiled, shared and analyzed much faster. Medical workers in Kenya conduct house-to-house visits doing HIV counseling and testing using ODK-equipped smartphones to track patients’ medical histories (accessed by using the phone to scan a bar code on a patient’s ID card) and upload geo-coded information directly to the health information system [9].

A variety of new data mining and visualization tools and technologies are being developed to provide enhanced views of large sets of aggregate data [10]. A promising new approach to real-time data collection for research and response involves mining streams of data as they are generated on the Internet. Several tools have been developed to gather distributed data via the web (or alternative data streams) and visualize it in real-time. InStedd’s Riff tool is an interactive decision-support environment that combines the power of virtual teams of human experts and advanced analytic, machine-learning and visualization services to allow its users to collaborate around streams of information in order to detect, characterize and respond rapidly to emerging events. During the H1N1 pandemic, Riff was used to mine Google searches related to flu symptoms, identify where those searches were coming from in the United States and map the locations, resulting in accurate prediction of actual outbreaks as they were developing and opportunity to develop prevention responses such as school closures [11].

Ushahidi, similar in purpose to Riff, supports gathering of distributed data from the web and other data streams. Developers throughout Africa and beyond are using Ushahidi to extract and map a variety of types of data including crisis response and recovery (Chile, Haiti) and medical and pharmaceutical stockouts (Kenya, Uganda, Malawi, Zambia) [12].

EpiVue, developed by the Center for Public Health Informatics, University of Washington, integrates open-source technologies to provide a geospatial visualization framework for public health data. Users can upload data sets in a variety of formats and visualize the data via Google map [13].

Zook et al. [14] outline the ways in which a variety of information technologies including crowdsourcing for online mapping were used in the Haiti relief effort. They demonstrate the potential of crowd-sourced online mapping and the potential for new avenues of interaction among physically distant places to plan prevention strategies.
Conclusion

Prevention informatics offers a useful paradigm for re-imagining health information systems and for harnessing the vast array of data, tools, technologies and systems to respond pro-actively to health challenges across the globe. Research opportunities for information scientists in prevention informatics abound. Critically needed are methods to optimize timely data exchange of information between (to and from) clinical and public health information systems in order to improve speed of prevention response for individual and community health. Citizen-generated information offers new means to detect disease threats and respond to disasters as well as offering communities of practice to support prevention response in resource-constrained environments. At the same time, with the availability of instant communications, we need to understand how to recognize and prepare for unexpected crowd reactions to threats. Improved vocabularies, thesauri and ontologies of concepts, coupled with visualization tools and GIS mapping, can provide new insights across disparate information resources and databases from information and data sources across the globe.

Acknowledgements

A special thank you to my public health mentors, Rear Admiral Dr. Patrick O’Carroll, Region 10 Health Administrator, Public Health Services, whose lectures on public health and the importance of prevention inspired me to think more broadly about health information systems design; to Dr. Mark Oberle who provided a comprehensive grounding in the vital work that public health professionals do even when they are on vacation; and to Drs. Roddy Stusser and Walter Curioso who contributed their perspectives on global issues in prevention informatics.

Resources Mentioned in the Article