Smart Infoware: Providing University Research Stakeholders Soft Power to Connect the Dots in Information Haystacks

Chi-Ren Shyu, Director, Informatics Institute; Shumaker Endowed Professor of Informatics, University of Missouri

While resources have been allocated to build computing infrastructure everywhere in the nation, the value of infoware to assist the university scientific communities has been underestimated, or even ignored. Those unsung heroes, who develop infoware to dig into the information haystacks and provide to the research community needle-sized up-to-date knowledge, deserve recognition, since without their innovative work scientific discovery would not be able to move quickly from bench to bedside in healthcare or move from greenhouse to dishes in agronomy and food production. Moreover, the organization and coordination of available infoware are needed to leverage regional talents to equip researchers with soft power as opposed to the hardware-based computing muscles.

Informaticians and Infoware

Informaticians often identify research problems and then model potential solutions in a mathematical or computational way in order to streamline the knowledge discovery process so that costly experiments will be minimized. I would like to use University of Missouri Informatics Institute (MUII) to demonstrate the development process, application, and potential of infoware developed by the informaticians. The history of MU informatics research began in the 1960s when Dr. Donald A. B. Lindberg, the Director of National Library of Medicine, pioneered the application of computer technology to healthcare. In 2012, there are 42 core faculty members from 14 departments and 7 colleges/schools who contribute to the curriculum development and research/outreach activities to support a doctoral program hosting 35 PhD

Figure 1. In the personalized medicine era, physicians will need to hash through complex information for accurate and personalized care.

students. The informatics community at MU continuously develops infoware to serve the worldwide research community in three major areas – bioinformatics, health informatics, and geoinformatics. This infoware has played significant roles in handling the ever- growing size of data in genomics, proteomics, biometrics, and imaging technologies. As shown in Figure 1, future healthcare will need to connect the dots from sensor data for homecare patients; understand various image modalities; recognize protein structure data for drug reactions;

Figure 2. Number of protein structures deposited in the Protein Data Bank (Picture from the PDB database).

and comprehend personalized sequence data. It becomes more and more challenging when the goal of "$1000 genome" will soon be reached. It is a known joke in the industry that "sequence a whole genome overnight and take months to analyze it." The bottleneck lies in the computational inability to process the data in each lab without shipping the data through ultra-high-speed network to a high performance computing facility

Bioinformatics

Figure 2 shows the growth trend of protein structures in the Protein Data Bank (PDB). Without smart infoware to analyze the data, the financial burden to purchase larger computer clusters becomes bigger and unmatchable to the growth of information. Examples of infoware developed by MU bioinformaticans include:

SoyKB (http://SoyKB.org): Soybean knowledge base is a comprehensive all-inclusive web resource for soybean research. It provides an infrastructure to handle the storage, integration, and analysis of the gene, genomics, EST, microarray, transcriptomics, proteomics, metabolomics, pathway and phenotype data.

ProteinDBS (http://ProteinDBS.rnet.missouri.edu): A Web server designed for efficient and accurate comparisons and searches of structurally similar proteins from a large-scale database. It provides worldwide users a speedy search engine for global-to-global and local-to-local searches of protein structures or substructures.

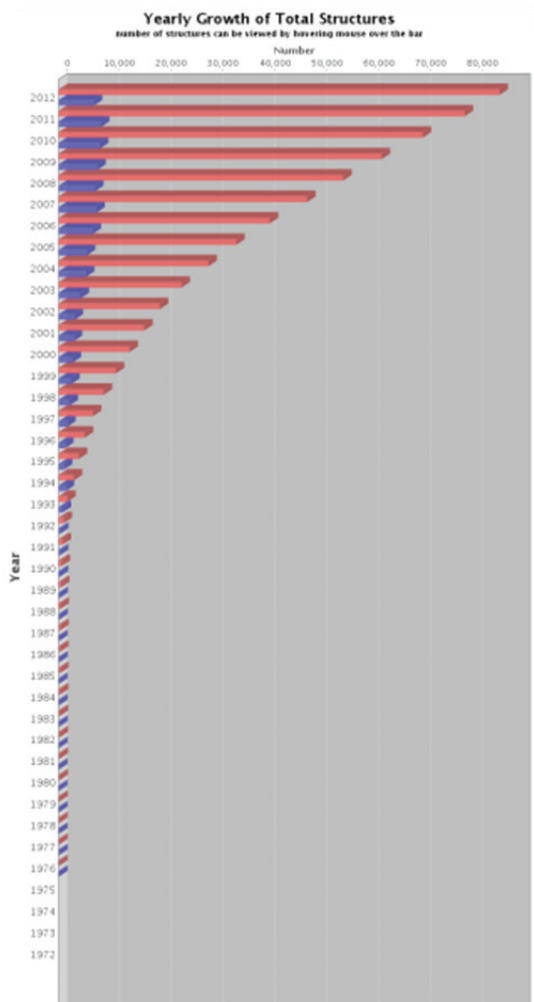DOMMINO (http://dommino.org): A comprehensive database of macromo-

lecular interactions that includes the interactions between protein domains, interdomain linkers, N- and C-terminal regions and protein peptides. This infoware provides a flexible search for interactions and visualization tools to navigate the search results.

PhenoDBS (http://PhenomicsWorld.org): A web-based infoware that provides the genomics community with a computational framework to study complex phenotypes using techniques such as visual content management, semantic modeling, knowledge sharing, and ontology customization.

All the infoware shares the same goal, which is to provide open-source tool access to the entire scientific community with speedy search from large-scale and complex data sets that normally cannot be organized and processed without high performance computing.

Health and Medical Informatics

Billions of dollars have been invested to drive innovations in health care technologies, such as the Nationwide Health Information Network (NHIN), to enable health information exchange with secured, simple, and seamless infrastructure. This provides physicians an environment with meaningful use of electronic health records, and Centers for Medicare and Medicaid Services (CMS) has the

Health Care Innovation Awards to implement the most compelling new ideas to improve the quality of the care as well as reduce the cost for citizens enrolled in Medicare, Medicaid, and Children's Health Insurance Program.

One example of infoware developed by MU health/medical informaticians is the informatics framework for the American Lymphedema Framework Project (http://ALFP.org). This project provides an operational cyber infrastructure that collects, organizes, and disseminates up-to-date lymphedema (LE) information. It includes the development of a cyber framework for inter-institutional lymphedema research activities; the integration of informatics tools to provide an on-line summary of concurrent LE studies; and the creation
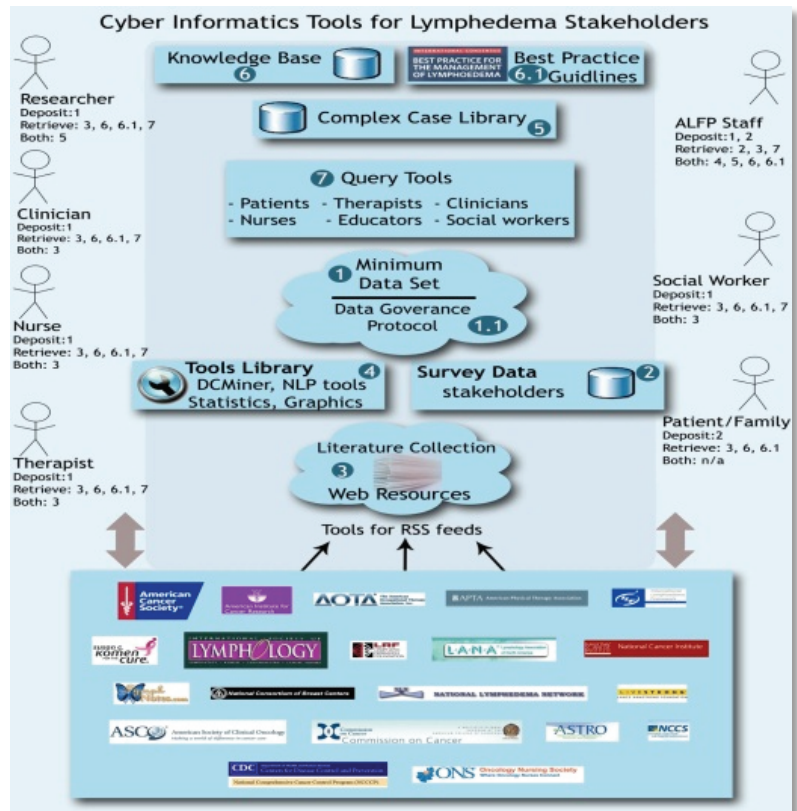

Figure 3. A medical informatics infrastructure to provide infoware to serve an international community in lymphedema.
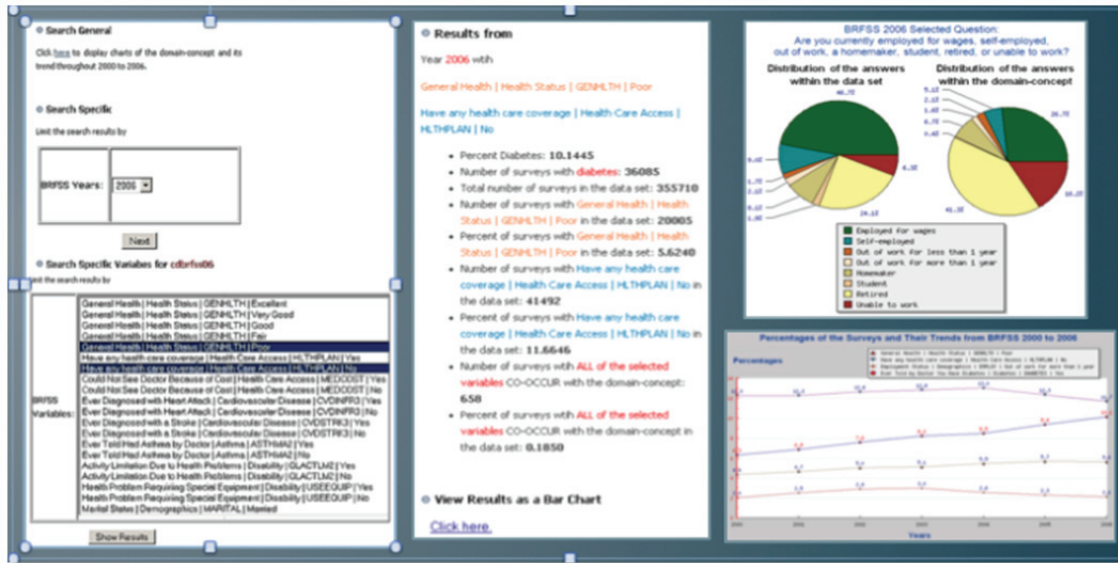
57

Figure 4. Infoware for clinicians to map a single patient's profile to high-risk populations for real-time alerts and disease management.

of a web portal for LE stakeholders with user-specific query methods. Infoware developed for this project have been targeted to serve users from all ranges of backgrounds, such as patients, therapists, physicians, researchers, etc. In addition, mobile apps, such as the iPhone App "Look4LE" as shown in Figure 3, are also developed to provide information access to patients.

Another infoware example in health and medical informatics is mapping tools for researchers and health professionals to connect a specific patient with public knowledge through public health information systems, such as cancer registry and Behavioral Risk Factor Surveillance System (BRFSS) data from the Centers for Disease Control and Prevention (CDC). Applying mining tools to extract clinically significant patterns for various demographic and geographic populations, researchers and clinicians can compare a patient's profile with the trends of healthcare related activities for high-risk patients from the public infor-

mation systems. This infoware can provide real-time alerts for patient's disease self-management and for clinicians to provide proactive cares.

Geoinformatics

More and more data are expected to be geo-coded when the number of data acquisition devices with embedded GPS functions increases. Geographic Information System (GIS) analytics can be used to provide the research community the ability to analyze geospatial information for decision-making. It has been postulated that GIS could provide distribution of available information resources in order to bridge the gap between at-risk patients and access to therapists and treatment centers for chronic diseases. GIS-based infoware is utilized to analyze patients and chronic disease resources as a geographical representation in order to identify associations between them. Additionally, we sought to provide information to inform policymakers and educators in order to assist with resource development, prioritiza-
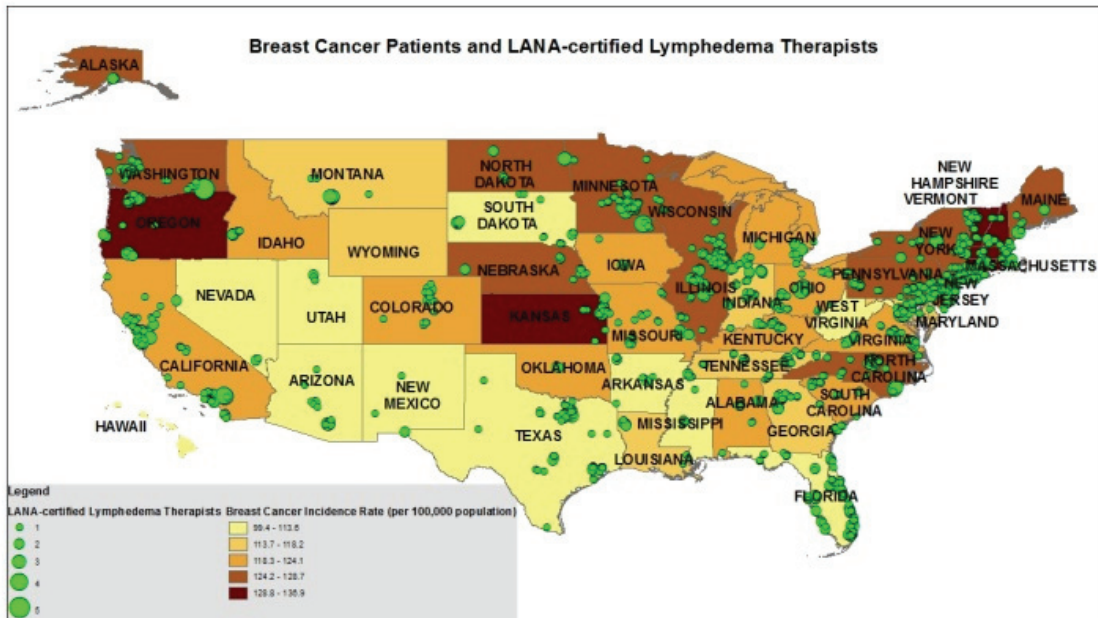
58

Figure **5**. A GIS-based view of available resources for breast cancer survivors who might suffer from certain chronic diseases.

tion, and allocation. Figure 5 shows available certified therapists for lymphedema patients.

Scenarios Involving Multiple Informatics Areas

One example of infoware that involves techniques from both bioinformatics and geoinformatics is the development of tools that can allow plant genetic researchers and breeders to search crop plant disease phenotypes in real-time using genetic, phenomic, and geospatial information. The technology has been tested with the maize community to study mutants and diseases
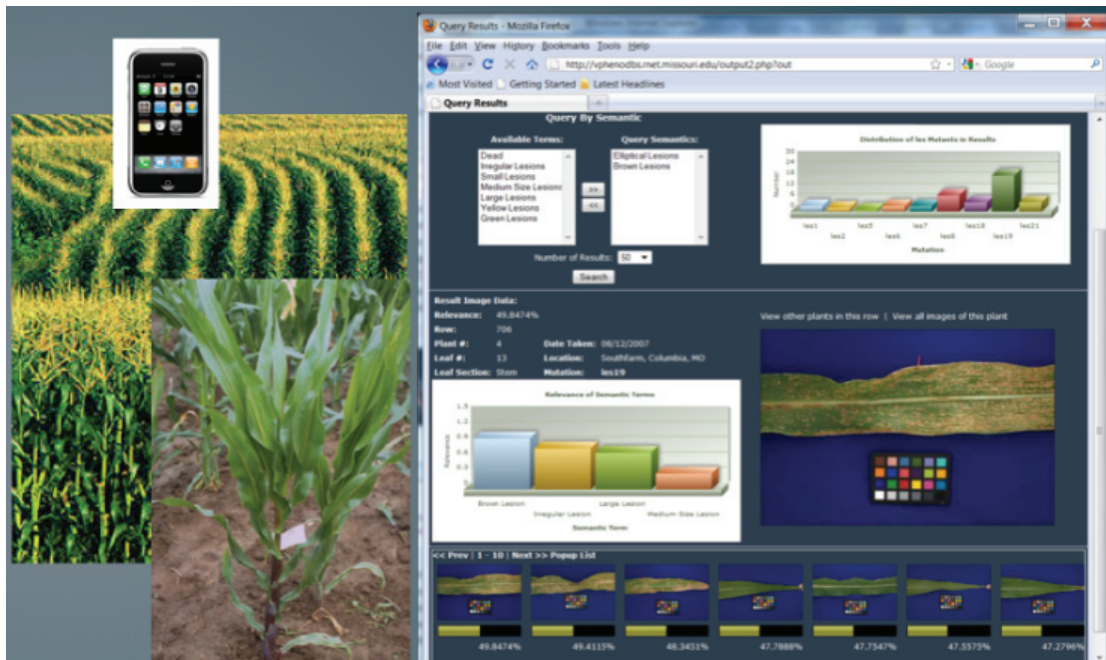

Figure 6**.** An advanced search engine to look for similar plant images for some diseases.

to streamline the process of plant morphology, ecology, and phytochemistry research. Such infoware is unique in the sense that it provides GIS-enabled query by phenotype images, query by semantics in mutant and disease descriptions, and other customized complex query methods to assist plant researchers studying the underlying effects of genetics and environmental factors on physiology. Figure 6 shows the advanced search tools that will enable future research in plant genetics by fostering cross-

necessary for infoware developers from the region to meet and put together an info-warehouse for tool sharing and education.

Moreover, a research social network which is searchable by university researchers and industry partners is also needed for the region. This linkage of researchers may consist of co-authored publications, collaborative proposals for extramural grants, student committee memberships, national/international committee services, etc. Challenges occur mainly on data collection
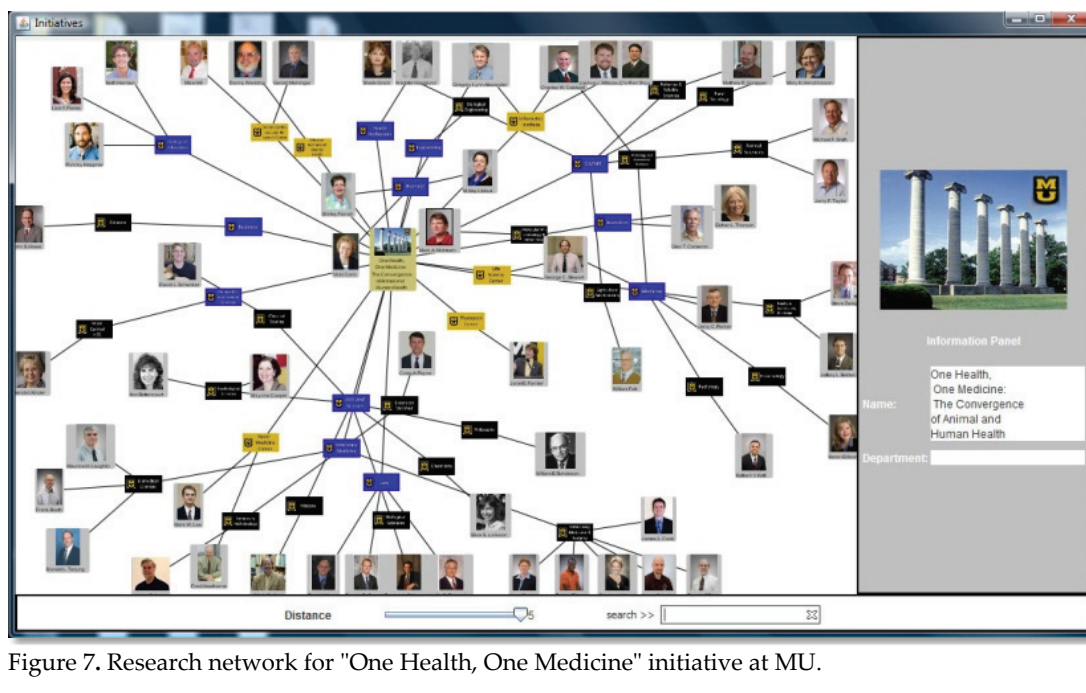

Figure 7. Research network for "One Health, One Medicine" initiative at MU.

institutional study of complex traits using unique infoware developed by informaticians.

Conclusions

Infoware has been developed independently by colleagues in Iowa, Kansas, Missouri, and Nebraska. However, most informaticians are unaware of these developments in their surrounding institutions. Thus it is unlikely to expect researchers in other fields to understand the regional talents that can greatly enhance their research using the existing infoware. Therefore, it is

and quality control when multiple institutions are involved for this talent knowledge base. Figure 7 illustrates an example of a research network for researchers who have collaboration records in the area of the One Health and One Medicine Initiative under the Mizzou Advantage program. To make the network valuable, an ideal infoware for such type of network should provide talent search as powerful as leading search engines, such as Google™, and recommend researchers as thoughtful as popular shopping sites, such as Amazon.com™.