

# Making use of electronic data: The National Healthcare Safety Network eSurveillance Initiative

Jonathan R. Edwards, MS, Daniel A. Pollock, MD, Benjamin A. Kupronis, MPH, Wenkai Li, MD, MAS, MS, James S. Tolson, BS, Kelly D. Peterson, BBA, Randy B. Mincey, MBA, and Teresa C. Horan, MPH  
Atlanta, Georgia

Efforts are underway at the Centers for Disease Control and Prevention to foster greater use of electronic data stored in health care application databases for surveillance of health care-associated infections and antimicrobial use and resistance. These efforts, referred to as the National Healthcare Safety Network (NHSN) eSurveillance Initiative, focus on standards-based solutions for conveying health care data and validation processes to confirm that the data received at the Centers for Disease Control and Prevention accurately reflect the data transmitted by health care facilities. Standard vehicles for data transmission, specifically Health Level Seven standards for electronic messages and structured documents, and standard vocabularies for representing microorganisms and other information needed for surveillance, are central features of the eSurveillance Initiative. Progress to date in this initiative is reviewed, and future project plans are outlined. Enhanced interoperability between health care and public health information systems is achievable for surveillance purposes, but major challenges must be overcome to realize the full benefits sought by the eSurveillance Initiative. (*Am J Infect Control* 2008;36:S21-6.)

Electronic data transfer between health care and public health information systems can yield important benefits for surveillance of health care-associated infections (HAIs) and antimicrobial use and resistance (AUR). More timely, efficient, comprehensive, and reliable HAI and AUR surveillance can be achieved by capitalizing on advances in health care information technology, medical informatics, and increasing availability of health care data in electronic form.<sup>1,2</sup> However, capturing and conveying HAI and AUR data electronically from existing hospital systems present many challenges, 3 of which are described in this paper.

First, electronic data must be transmitted in a standardized format that allows for flexible and efficient data storage in human readable form and machine processability of the data at the receiving system.

This standard format is optimally based on a well-designed information model in which the relationship between data elements can be represented and maintained as data elements are combined, added, or removed from the model. This standard way of packaging electronic data is considered the vehicle. The second challenge in capturing electronic data is achieving standardization in how the data are represented. For example, pathogen information often is stored in hospital laboratory information systems using local codes for specific microorganisms. These codes convey meaning to those familiar with their use. However, variability in representing the same pathogens in different laboratory information systems presents a problem when aggregating data from multiple health care facilities. The solution is standard terminology and codes for the same concepts, ie, standard vocabularies for purposes of concept representation and communication. Although standard vocabularies are available for many concepts that are central to HAI and AUR surveillance, these vocabularies are far from ubiquitously implemented in health care information systems. Up-front investment in vocabulary mapping from local to standard codes is a common prerequisite to enable interoperability between sending and receiving information systems. The third challenge is maintaining or improving the quality of the data content transmitted between systems. Data can be conveyed electronically using a standard vehicle and standard vocabularies, but the end results will fall short with respect to quality if the data received does not accurately reflect what was originally in the sending system. This challenge highlights the need for ongoing validation

From the Division of Healthcare Quality Promotion, National Center for Preparedness, Detection and Control of Infectious Diseases, Centers for Disease Control and Prevention, Public Health Service, US Department of Health and Human Services, Atlanta, GA.

Address correspondence to Jonathan R. Edwards, MS, Centers for Disease Control and Prevention, Division of Healthcare Quality Promotion, 1600 Clifton Road, NE, MS-A24, Atlanta, GA 30333. E-mail: [jredwards@cdc.gov](mailto:jredwards@cdc.gov).

The findings and conclusions of the report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

0196-6553/\$34.00

Copyright © 2008 by the Association for Professionals in Infection Control and Epidemiology, Inc.

doi:10.1016/j.ajic.2007.07.007

and targeted improvements in data quality in both sending and receiving systems.

Responding to these 3 challenges has been a major focus of the eSurveillance Initiative of the National Healthcare Safety Network (NHSN), a surveillance system developed and maintained by the Division of Healthcare Quality Promotion (DHQP), National Center for Prevention, Detection, and Control of Infectious Diseases (NCPDCID), Centers for Disease Control and Prevention (CDC), and its surveillance partners. In this report, we provide as background an overview of the NHSN, the need for accessing electronic data for surveillance, and detail the NHSN eSurveillance Initiative. We also summarize the current activities and outline future plans in addressing the challenges of capturing and conveying health care data in electronic form for purposes of HAI and AUR surveillance.

## BACKGROUND

Public health surveillance is a core activity at the CDC and an integral part of its prevention efforts across a wide range of programs. The NHSN is the system the CDC uses for national surveillance of patient and health care personnel safety and antimicrobial resistance. The NHSN was deployed in October 2005 and is the successor system to 3 CDC legacy systems: National Nosocomial Infections Surveillance (NNIS) system, National Surveillance System for Healthcare Workers (NaSH), and Dialysis Surveillance Network. From the inception of the NNIS system in 1970, the CDC's primary goals in HAI surveillance have been to describe the epidemiology of HAI, provide national-level HAI rates for hospitals to use for comparative purposes, and promote methodologically sound surveillance in hospitals.<sup>3</sup> In more recent years, as concerns about antimicrobial resistance have mounted, surveillance of resistance has been added as a primary goal for the NHSN.

For HAI events and antimicrobial resistance, surveillance data are collected through the NHSN that enable rate calculations. In the case of antimicrobial resistance, the data suggest that rates are increasing each year and that new patterns of resistance are developing.<sup>4,5</sup> Infections with resistant organisms cause economic, risk management, and other problems for health care institutions. Many of these organisms persist or recur in hospitals, especially in intensive care units (ICUs). Control of resistant organisms depends on accurate, comprehensive, and timely data to help guide changes in patient care practices. Rate data are needed to track trends over time and across care settings and to evaluate the impact of interventions.

Conventional methods of data collection for HAI and antimicrobial resistance surveillance involve a series of

procedural steps that are time-consuming, costly, and potentially error prone.<sup>6</sup> These methods often include an initial step in which patient data are manually abstracted from clinical and laboratory records, which in some cases provide data in electronic form. In a subsequent step, data are manually entered into special purpose data collection forms of a surveillance system for storage, analysis, and data sharing. An alternative to this conventional approach is to capture data that are already available in electronic form and automate transmission of those data to a receiving system, such as the NHSN. This alternative approach to data collection and transmission, which we refer to as *eSurveillance*, obviates the need for manual data abstraction from electronic records and reentry of those same data into a surveillance system. The initial goal of the NHSN eSurveillance Initiative is enabling electronic capture and transmission of data needed for surveillance of AUR in health care settings.

In terms of information technology infrastructure and functionality, the NHSN system is a centrally developed and maintained technology platform composed of multiple products used to collect and disseminate meaningful health care-associated data. One of the main products developed to support the NHSN system is the NHSN Web-based application, which allows users to enter, import, validate, and analyze their health care-associated data. Health care facilities that participate in the NHSN use their own data for internal, clinical performance measurement and quality improvement purposes. As a result, NHSN participants are motivated to collect data that are timely, accurate, and directly pertinent to their own institutional needs. Important secondary benefits accrue when the CDC aggregates facility-level data for national surveillance of HAI and antibiotic resistance trends. The pivotal linkage between the motivation of NHSN participants and NHSN's capacity for public health surveillance places a premium on assuring first and foremost that the Web-based application meets NHSN user needs and that application enhancements are user oriented. This linkage also provides an important impetus for the efforts underway through the NHSN eSurveillance Initiative to streamline use of electronic data for facility- and national-level surveillance.

The eSurveillance Initiative is a research and development component of the NHSN. The Initiative is designed to produce technical solutions for the NHSN that draw on prevailing and rapidly emerging health care information technology standards and that enable increased use of electronic data for HAI and AUR surveillance. The work for this project began in 2002 with development of specifications for how data will be packaged and conveyed (ie, vehicle) for AUR surveillance. The prevailing industry standard for clinical and

laboratory data exchanges is provided by Health Level Seven (HL7), an American National Standards Institute-accredited, standards development organization that is best known for producing and maintaining electronic message standards.<sup>7</sup> The eSurveillance team selected the HL7 version 3 (V3) message standard for the project in anticipation of industry transition from HL7 V2 to V3 messages in data exchanges that involve the 3 source hospital information systems needed for AUR surveillance: laboratory systems for microbiology susceptibility laboratory test results; pharmacy systems for antimicrobial use; and admission/discharge/transfer (ADT) systems for patient location and other basic demographic and hospital use characteristics.

The CDC eSurveillance team, working with a clinical information system vendor and an HL7 message implementation consultant, pilot tested the HL7 V3 messages in prestandard form to develop V3 message implementation guidance and the technical infrastructure to support V3 messages between sending and receiving systems. The decision to use V3 was supported by several factors. First, V3 messages are based on an object-oriented development methodology and the HL7 Reference Information Model (RIM), which is a comprehensive, high-level information model designed to guide message development and ensure that oversights and crossed assumptions are addressed early in the development process. Second, V3 provides a robust set of data types and explicit rules for how controlled vocabularies are to be used in messages. Third, V3 specifies messaging using the extensible markup language (XML), a World Wide Web Consortium (W3C)-recommended data format for exchanging a wide variety of data on the Web and elsewhere.

In October 2003, the eSurveillance team implemented secure transmission of microbiology susceptibility laboratory test result data using an HL7 V3 message sent from a single pilot health care facility to the CDC, at which the data were parsed into a database. This message implementation was facilitated by a clinical informatics software vendor with extensive experience reading and writing HL7 messages. Since the initial message implementation, messaging between pilot sites and the CDC has been extended to transmission and parsing of V3 pharmacy and ADT messages. As byproducts of these efforts, the eSurveillance team has interacted with the HL7 organization and contributed to development of V3 standards, iteratively refined message specifications for use in the eSurveillance project, and developed implementation guidance that can be used in data transmissions from other health care facilities to the CDC. The number of eSurveillance pilot sites has increased to 5 health care facilities: 2 with vendor facilitation and the other 3 using a software tool called the Data Transform Engine (DTE) developed

at the CDC. The DTE provides facilities that have created data stores containing the necessary data with the ability to generate HL7 V3 messages for secure transmission to CDC. Among these 5 participating pilot sites, 101,215 messages (40,171 microbiology, 58,981 pharmacy, and 2063 ADT) had been transmitted as of April 2007 and subsequently parsed.

The benefits of receiving large numbers of messages would be limited if not for the accompanying efforts to identify, develop, and use standard vocabulary for all the coded data elements in the 3 message types. The vocabulary for representing microorganisms often differs between health care facilities. Although local terminology and codes carry meaning within a single institution, they may have little semantic viability externally unless they are mapped to a standard vocabulary that supports aggregation across multiple facilities. For example, a given health care facility may use the code ENT to represent *Enterococcus* species; however, another facility may use it to represent *Enterobacter* species. Clearly, problems would arise when aggregating microbiology results from these 2 facilities. These vocabulary problems can be amplified quickly when expanding the scope, in this case, beyond a single genus or by simply increasing the number of participating facilities.

Another important issue when considering common terminologies is the need to maintain any inherent hierarchy among code values. In the example above, the first facility represents *Enterococcus* species using ENT while representing *Enterococcus faecalis* as FC. Although a more logical code for *Enterococcus faecalis* would be ENTFC, there may still be no real way to relate ENTFC to ENT in a hierarchical manner. For instance, what if the code ENTFC represented either *Enterococcus faecalis* or *Enterococcus faecium* without specificity. To attempt to resolve these and other problems with local terminologies, standard vocabularies have been developed. Table 1 displays a list of standard vocabularies used in AUR messages in the eSurveillance Initiative.

## CURRENT ACTIVITIES

The 3 HL7 V3 messages used for AUR surveillance provide a flexible, testable, and object-oriented approach. The 3 message types, namely, microbiology susceptibility test results, pharmacy, and ADT, are generated with varying degrees of frequency. Because of the potential high frequency of either pharmacy or ADT transactions in the health care setting, existing information technology infrastructure could be strained beyond capacity. This problem presented the need to identify a more efficient method for limiting data transmission frequency to reduce the load on hospital

**Table I.** Standard vocabularies used in AUR messages of the NHSN eSurveillance Initiative

| Data element                        | Standard vocabulary   |
|-------------------------------------|-----------------------|
| Gender                              | HL7 V3                |
| Race                                | HL7 V3                |
| Ethnicity                           | HL7 V3                |
| Discharge status                    | SNOMED CT             |
| Discharge diagnosis                 | ICD-9-CM              |
| Microorganism identification status | HL7 V3                |
| Microorganism                       | SNOMED CT             |
| Microbiology test method            | LOINC-HL7             |
| Susceptibility interpretation       | HL7 V3                |
| Microbiology result unit of measure | HL7 V3                |
| Body site                           | SNOMED CT             |
| Specimen type                       | SNOMED CT             |
| Specimen collection method          | SNOMED CT             |
| Health care delivery location       | NHSN location codes   |
| Drug ingredient                     | RxNorm                |
| Drug form                           | RxNorm                |
| Route of administration             | NCI thesaurus-FDA SPL |
| Dose frequency                      | HL7 V3 (GTS)          |
| Unit of dose quantity               | NCI thesaurus-FDA SPL |

ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; LOINC, Logical Observation Identifiers Names and Codes; NCI, National Cancer Institute; FDA SPL, Food and Drug Administration Structure Product Labeling.

computing resources. HL7 provides support for more efficient storage of high-volume transaction data through the use of a batch protocol. The HL7 V3 batch protocol allows multiple results linked to a specific entity (eg, patient) and multiple entities to be packaged into a single message without repeating header information. In this case, header information records provide details about the entity level. Avoiding the repetition of this entity level information enables more efficient storage, which reduces physical size requirements.

The volume of messages that have been parsed is increasing. However, this accumulation of data represents many months of transactions calls for further analysis before the data can be transformed into usable information. An important prerequisite is confirming that the data received at the CDC are those that were intended to be sent from the source health care facility. The ability to validate these messaged data is best enabled through the formulation of acceptable metrics. Simple counting schemes allow for easy comparison between pre- and post-transmission. For instance, microbiology results can be compared by counting the number of resistant as well as total tested isolates for a given microorganism-drug combination and then calculating antimicrobial resistance rates. Comparisons of these rates can easily be performed to assess their level of agreement, with the goal being equality. In the case of ADT data, comparisons can be performed by counting days of patient stay through the use of a common computational algorithm. Additionally, the simple act

of analyzing data provides for the opportunity to spot inconsistencies in the data or unexpected coding. This validation effort has allowed for not only the assessment of accuracy but also the consistency of rendered data. If the source data were somehow changed, the validation process would need to detect it.

Vocabulary standardization and mapping are critical steps in the process of creating usable information for surveillance of health care-associated events. This project has benefited greatly from the work of vocabulary specialists who have created highly flexible and well-maintained vocabularies. For many coded data elements, existing vocabularies can be adopted directly or modified with some enhancements to fulfill the requirements of the data being transmitted in this project. However, a controlled vocabulary was lacking for one salient data element: the locations within health care facilities in which patient care is delivered. For example, a medical critical care unit needs to be distinguished from a general medical ward. This is particularly important for data accumulation in the NHSN system because this distinction is important for proper stratification of aggregated measures such as rates of HAI or antimicrobial resistance. To address this need, efforts are underway to develop, promote, and mature a standard vocabulary for patient care location that harmonizes codes used in the NHSN with those used in HL7 for patient seen location.

Creation of these 3 HL7 V3 message specifications has been well complemented by the participation of the 5 pilot health care facilities. These institutions have enabled guided implementation of these specifications, which is critical to their refinement. One of the longer term goals of this project is to be able to publish these specifications for use by other health care institutions and health care system vendors. To promote the use of these 3 message types, special technical documents are being developed. These documents, called *Implementation Guides*, provide explicit details on how to create these types of messages as well as how to map and populate the data elements contained within the messages. Developing Implementation Guides will greatly enhance the acceptance of these message specifications as standard.

Recently, 2 new efforts have been added to the NHSN eSurveillance Initiative: (1) collaboration with CDC's National Center for Public Health Informatics on use of the Biosense technical infrastructure for transmission of HL7 version 2.5 messages conveying microbiology results, pharmacy, and ADT data and (2) collaboration with HL7 and a group of commercial software vendors to develop HL7 Clinical Document Architecture (CDA) specifications and implementation guidance for HAI reporting from vendor infection control software to the NHSN. This CDA development work

has been prompted in part by legislative mandates in an increasing number of states that require hospitals to report HAI data to a state agency and the decision in some states to use the NHSN as the technical infrastructure for hospital reporting.<sup>8,9</sup> The CDA solution under development is designed to enable HAI data already entered into vendor software to be reported to the NHSN without requiring duplicate data entry. The CDA standard provides semantic interoperability for reporting HAI data among the growing network of public health partners.

The precise requirements of the various state mandates vary; however, one main theme is the reporting of HAI rates. This requires the need to transmit HAIs and their respective denominator data. To enable a health care institution or health care information system vendor to transmit HAIs and denominator data electronically, CDA specifications and implementation guidance is needed for numerator and denominator data collection forms. As a priority, CDA document specifications are being developed to facilitate electronic transmission of bloodstream and surgical site infections along with their denominators of device-day summary and procedure data, respectively.

## FUTURE PLANS

Converting these electronic messaged data into usable surveillance information is of the utmost importance and should be accomplished over the next 5 years. One of the original objectives of capturing electronic microbiologic, pharmacologic, and ADT data was to populate the data required on the manual data collection forms containing summarized microbiology susceptibility results and antimicrobial usage quantities. At the time, these data were collected following the AUR component of the NNIS system. Because nearly identical data collection forms exist within the Medications-associated module of the NHSN, these electronic data will supplant the manually collected data using the NHSN. Populating these data electronically will allow a facility to participate in the Medications-associated module of the NHSN. Furthermore, it will permit the facility to perform statistical comparisons with aggregate data.

The potential uses of these data exceed the basic population of manual data entry forms. The capability to populate specific susceptibility results for HAIs required by the Patient Safety Component protocols of the NHSN also could be realized, which could reduce the HAI collection burden by an estimated 30%.

Several new areas of HAI research could be advanced with these microbiology, pharmacy, and ADT data. HAI detection algorithms could be evaluated for agreement with standard manually detected methods.

The CDC is collaborating with the CDC Prevention Epicenter Program investigators to assess whether automated bloodstream infection rates generated from electronic data are acceptable surrogates for rates calculated using infection control professionals' surveillance data.<sup>10</sup> Other areas of investigation by the Epicenters Program include using electronic data to improve surveillance for surgical site infection and *Clostridium difficile*-associated disease and generating measures of antimicrobial utilization that could help optimize antimicrobial use. Electronic data could also be used to follow trends of emerging resistance allowing a more timely response to new threats. In addition, new measures associated with surveillance efforts could be used to track, control, and reduce the spread of multidrug-resistant organisms.

The validity and accuracy of HAI detection algorithms must be assessed. This requires the use of a "gold standard" to which all algorithms can be compared. There needs to be a standard methodology for assessing sensitivity and specificity among measures. Substantial efforts and collaboration among stakeholders conducting HAI event detection research are critical for valid accuracy evaluation.

The 3 HL7 V3 message specifications will be promoted by 3 main activities. As more facilities send these messages to the CDC, the utility of rendered data will increase, which in turn will increase viability of these specifications. Next, the Implementation Guides will support their use. Last, the message specifications will be promoted through the HL7 balloting process, which will ultimately result in being accepted as messaging standards for these types of data.

Some of the standard vocabulary for the data elements represented within the 3 message types will require further maturation. In some cases, sufficient development time has been invested to represent hierarchical concepts and terms in a well-designed manner. For instance, to depict organism concepts, one of the nationally recommended terminologies in the health care domain is the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), which is maintained by the College of American Pathologists.

In a few cases, there are concepts and terms for which considerable effort will be needed to produce and maintain valid and usable taxonomies. Taxonomies should enable a consistent way of capturing, sharing, and aggregating health care data. As mentioned, one of the important tasks of this project is to develop a robust reference terminology to depict and describe patient care or health care delivery locations. This effort will require development of a hierarchical design and a list of definable attributes and values that enhances the reference of these important terms.

Reporting of HAI data to public health partners using CDA is a promising area of standards development. Widespread adoption of CDA will require an appreciation for balancing the imposition of minimal constraints on local practices with the benefits of enhanced data interchange among all exchanging parties especially at state and federal levels. Templates will need to be created to support the potential list of HAIs to be exchanged. Partners implementing this standard will need to undertake the potentially complex task of validation. Vocabulary access and distribution are critical to assisting local partners in their implementation of CDA. All of the aforementioned are crucial to maturing the CDA specifications for HAIs, procedures, and summary data.

## CONCLUSION

A standards-based approach is imperative to the success of a scalable national surveillance system requiring transmission of electronic data. Gone are the days of dumping data into flat files and sending to a destination on floppy disk. Although this may be a relatively easy method to transfer data between a single set of partners, the task becomes exponentially more complex when multiple partners are involved. Even in the latter case, participants may be able to agree on a defined format for transmitting data, yet it may not be easy to verify the correct use of the format without a mechanism to test whether or not the correct structure was implemented.

Source systems or public health partners need to be able to exchange data in a well-defined manner. This will enhance the interoperability of data interchange between participating entities, raising the quality and utility of the information being exchanged.

Validation of the exchanged data is critical to the success of electronic surveillance. There is a strong need for a standardized, interoperable, and validated approach that best serves the needs of clinical performance measurement and public health surveillance,

and the NHSN eSurveillance Initiative exists to address this need.

The authors thank Mead Walker for his expert contributions as a HL7 message implementation consultant and to Jonathan Olson, Brian Bray, and Rob Hausam, employed by TheraDoc Inc., a clinical information system vendor, who have supported this research and development work, and the 5 health care facilities that have actively provided these important electronic surveillance data. The authors also thank Alschuler Associates, a consulting firm with expertise in HL7 CDA, and a group of clinical information system vendors for their work in developing and pilot testing CDA R2 implementation guidance for HAI reporting to the NHSN. The 6 vendors that participated in a pilot are Cardinal Health Med Mined Services, EpiQuest, ICPA, Premier Inc, TheraDoc Inc, and Vecna.

## References

1. Trick WE, Zagorski BM, Tokars JI, Vernon MO, Welbel SF, Wisniewski MF, et al. Computer algorithms to detect bloodstream infections. *Emerg Inf Dis* 2004;10:1612-20.
2. Tokars JI, Richards C, Andrus M, Klevens M, Curtis A, Horan T, et al. The changing face of surveillance for health care-associated infections. *Clin Infect Dis* 2004;39:1347-52.
3. Richards C, Emori TG, Edwards J, Fridkin S, Tolson J, Gaynes R, and the National Nosocomial Infections Surveillance System. Characteristics of hospitals and infection control professionals participating in the National Nosocomial Infection Surveillance System 1999. *Am J Infect Control* 2001;29:400-3.
4. Klevens RM, Edwards JR, Tenover FC, McDonald LC, Horan T, Gaynes R, and the National Nosocomial Infections Surveillance (NNIS) System Hospitals. Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA) in intensive care units in US hospitals, 1992-2003. *Clin Infect Dis* 2006;42:389-91.
5. Fridkin SK, Kaufman D, Edwards JR, Shetty S, Horan T. Changing incidence of *Candida* bloodstream infections among NICU patients in the United States: 1995-2004. *Pediatrics* 2006;117:1680-7.
6. Horan TC, Gaynes RP. Surveillance of nosocomial infections. In: Mayhall CG, editor. *Hospital epidemiology and infection control*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2004. p. 1659-702.
7. Health Level Seven. Available at: <http://www.hl7.org>. Accessed November 14, 2007.
8. McKibben L, Horan TC, Tokars JI, Fowler G, Cardo DM, Pearson ML, et al. Guidance on public reporting of healthcare-associated infections: recommendations of the Healthcare Infection Control Practices Advisory Committee. *Am J Infect Control* 2005;33:217-26.
9. Wong ES, Rupp ME, Mermel L, Perl TM, Bradley S, Ramsey KM, et al. Public disclosure of HAIs: the role of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol* 2005;26:210-2.
10. Prevention Epicenter Program. Available at: <http://www.cdc.gov/ncidod/dhqp/epicenter.html>. Accessed November 14, 2007.