Clinical Microbiology and Infection 27 (2021) S29–S39

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Original Article

Information technology aspects of large-scale implementation of automated surveillance of healthcare-associated infections☆

Michael Behnke [1,*], John Karlsson Valik [2], Sophie Gubbels [3], Daniel Teixeira [4], Brian Kristensen [5], Mohamed Abbas [4], Stephanie M. van Rooden [6,7], Petra Gastmeier [1], Maaike S.M. van Mourik [8] on behalf of the PRAISE network[*]

[1)] *National Reference Center for Surveillance of Nosocomial Infections, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Hygiene and Environmental Medicine, Berlin, Germany*
[2)] *Department of Medicine Solna, Division of Infectious Diseases, Karolinska Institutet and Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden*
[3)] *Data Integration and Analysis Secretariat, Statens Serum Institut, Copenhagen, Denmark*
[4)] *Infection Control Programme, Geneva University Hospitals, Geneva, Switzerland*
[5)] *Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Copenhagen, Denmark*
[6)] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands*
[7)] *Centre for Infectious Disease Epidemiology and Surveillance National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands*
[8)] *Department of Medical Microbiology and Infection Control, University Medical Center Utrecht, Utrecht, the Netherlands*

ARTICLE INFO

Article history:
Received 17 November 2020
Received in revised form
24 February 2021
Accepted 25 February 2021

Editor: L. Leibovici

Keywords:
Automated
Bloodstream infection
Data
Digital infection control
Electronic HAI surveillance
Electronic health record
Healthcare-associated infection
Quality
Surgical site infection
Surveillance

ABSTRACT

Introduction: Healthcare-associated infections (HAI) are a major public health concern. Monitoring of HAI rates, with feedback, is a core component of infection prevention and control programmes. Digitalization of healthcare data has created novel opportunities for automating the HAI surveillance process to varying degrees. However, methods are not standardized and vary widely between different healthcare facilities. Most current automated surveillance (AS) systems have been confined to local settings, and practical guidance on how to implement large-scale AS is needed.
Methods: This document was written by a task force formed in March 2019 within the PRAISE network (Providing a Roadmap for Automated Infection Surveillance in Europe), gathering experts in HAI surveillance from ten European countries.
Results: The document provides an overview of the key e-health aspects of implementing an AS system of HAI in a clinical environment to support both the infection prevention and control team and information technology (IT) departments. The focus is on understanding the basic principles of storage and structure of healthcare data, as well as the general organization of IT infrastructure in surveillance networks and participating healthcare facilities. The fundamentals of data standardization, interoperability and algorithms in relation to HAI surveillance are covered. Finally, technical aspects and practical examples of accessing, storing and sharing healthcare data within a HAI surveillance network, as well as maintenance and quality control of such a system, are discussed.
Conclusions: With the guidance given in this document, along with the PRAISE roadmap and governance documents, readers will find comprehensive support to implement large-scale AS in a surveillance network. **Michael Behnke, Clin Microbiol Infect 2021;27:S29**

☆ This article is part of a supplement entitled Implementing Automated Surveillance of Healthcare-Associated Infections (HAI) sponsored by the PRAISE network (supported under 7th transnational call within the Joint Programming Initiative on Antimicrobial Resistance(JPIAMR), Network Call on Surveillance (2018) and funded by ZonMw (grant number 549007001)), the COMBACTE MAGNET EPI-Net project (funded by the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115523 | 115620 | 115737 | 777362 resources of which are composed of financial contribution from the European Union Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution) and Charité University Hospital.
* **Corresponding author**: Michael Behnke, National Reference Center for Surveillance of Nosocomial Infections, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Hygiene and Environmental Medicine, Hindenburgdamm 27, 12203, Berlin, Germany.
E-mail address: michael.behnke@charite.de (M. Behnke).
* Members of the PRAISE network are listed in the Acknowledgements.

https://doi.org/10.1016/j.cmi.2021.02.027
1198-743X/

Introduction

Healthcare-associated infections (HAI) are a major public health concern and a leading cause of morbidity, mortality and economic costs [1–3]. Monitoring the rates of HAI, with feedback of HAI rates to healthcare workers and stakeholders, is central to guide infection prevention and control (IPC) interventions in healthcare facilities [4]. Comparing HAI rates is important not only within a single facility but also in a wider context [5], and participation in a HAI surveillance network has been shown to result in decreasing rates of HAI [6]. Traditionally, HAI surveillance has been based on manual case-finding techniques, which are both labour intensive and subject to interrater variability bias [7,8]. Limitations in resources, which are not uncommon, restrict the size of the intended surveillance population when ideally all patients receiving healthcare ought to be surveyed. Increasing efforts to digitalize healthcare data have created novel opportunities for automating the surveillance process and thereby increasing efficiency and standardizing surveillance methods. However, to date, automated surveillance (AS) methods have been mainly developed by individual healthcare facilities and vary widely between them [9,10]. Practical guidance on how to implement large-scale AS and increase standardization is much needed. This information technology (IT) document accompanies the PRAISE network roadmap [11]. An article regarding the governance of AS also appears in this issue [12].

Our overall objective is to provide an overview of the key IT aspects of implementing an AS system of HAI in a clinical environment to support both project leaders and the IPC team as well as IT departments and to enhance collaboration in the development of large-scale AS. This article does not aim to cover all possible IT aspects of AS in depth; rather, it provides practical information and advice on how healthcare data is stored, how IT infrastructure is generally organized and how to succeed with AS in this setting.

The development and maintenance of AS systems requires access to a multidisciplinary team of epidemiologists, IPC nurses, clinical physicians, hospital IT staff, software developers, statisticians and project leaders. The intended audience for this document is primarily the infection prevention team, including the medical informatics and data analysts on that team, at the facility and national level. Also, the IT staff of healthcare facilities can use this document for an overview about the topic, although it could serve as a source of information for all readers with a specific interest in healthcare informatics in the context of HAI surveillance.

Methods

This IT document was written by a task force within the PRAISE network [11]. The group was formed during a workshop organized by the network in March 2019, which gathered 30 experts working in the field of HAI surveillance from ten European countries. The participants assigned to the IT task force all had experience or expert knowledge in healthcare informatics or practical aspects of developing and/or implementing automated HAI surveillance. After the workshop, members of the task force attended regular iterative teleconference meetings, and a draft of the document was presented at the second PRAISE network workshop in February 2020. Discussions and exercises performed by members of the entire PRAISE network, as well as continuous dialogue within the IT task force, were used to establish the document's outline and content. The process was supervised and supported by the network's core group. The document was critically reviewed for structure and content by all members of the PRAISE network, as well as by external expert reviewers within the field of HAI surveillance and healthcare informatics.

Scope and definitions

The focus is on automated HAI incidence surveillance for the purpose of comparison, prevention and quality improvement initiatives within surveillance networks. AS is defined as any form of surveillance where (parts of) the manual assessment are replaced by an automated process. This is often achieved by algorithms that use source data extracted from data routinely documented in electronic health records (EHR) during care (so-called routine-care data). Surveillance of HAI can be automated to varying degrees, and AS in this document refers to both semiautomated and fully automated surveillance.

In fully automated surveillance, the classification of HAI state data (numerator) and denominator data collection are performed by algorithms without any human interpretation, while semiautomated surveillance combines algorithms that retrospectively select high-probability episodes for manual chart review to ascertain HAI state (Box 1). Many approaches are possible to implement AS in a surveillance network, two of which are discussed and form the basis of the roadmap and supporting

Box 1
Examples of automated surveillance of healthcare-associated infections (HAI).

Fully automated surveillance of bloodstream infections (BSI)

For BSI, numerator data are based on microbiologic culture results of blood [13]. This includes metainformation such as sampling date, specimen cultured organisms, number of positive specimens for each cultured organism and for some cases also antimicrobial susceptibility testing results. Hospital-onset bloodstream infection (HOB) is defined as any positive blood culture obtained >48 hours after admission until discharge [14]. This does not require manual chart review of specific symptoms, making it well suited for automatic classification. The numerator of BSI (classification of HAI state) of BSI is usually linked to denominator data, such as number of days the patients were admitted to the hospital or specific wards (inpatient days). For central vascular catheter (CVC)-related BSI, denominator data on the number of days the patients have their CVCs (CVC days) needs to be extracted.

Semiautomated surveillance of surgical site infections (SSI)

The numerator in SSI surveillance is the occurrence of SSI along with details of the infection (e.g. date, depth, causative pathogen). Establishing the diagnosis of SSI usually requires clinical interpretation of the patient's condition using chart review. An example of a semiautomated solution for SSI is a classification algorithm applied to all patients at 120 days after prespecified surgical interventions to stratify according to a high or low probability of having developed an SSI [15]. Indicators of SSI are collected from routine-care data and dichotomized into subjects with a high probability of SSI (e.g. meeting a prespecified number of indicators), which are flagged for chart review. Every SSI is allocated to a specific surgical procedure. The denominator for SSI is based on the aggregated number of surgical procedures, making it easy to collect from structured data.

documents: centrally implemented AS and locally implemented AS (Fig. 1). The main distinctions are based on where the responsibilities for implementing surveillance and developing the algorithms are located. An elaboration of both approaches can be found in the roadmap [11].

A requirement for incidence surveillance is the classification of both HAI state data and calculation of denominator data. The HAI state data are a binary classification reflecting the presence or absence of HAI for each patient, patient episode or hospital episode, depending on the population definition, while denominator data are based on aggregated data collected from the entire population under surveillance. In HAI surveillance, denominator data are crucial for comparing infection rates in the local healthcare facility with other hospitals within or outside of the surveillance network. Together, the individual-level HAI state data and the denominator form the HAI surveillance result (Box 1).

Although HAI surveillance can be implemented as part of a research project, this document is concentrated on implementation with the purpose of improving quality of care for patients in the clinical setting. Relevant terminology and abbreviations are defined in the Glossary. For organizational, governance and legal aspects of HAI surveillance and data sharing, we refer to the discussion of governance aspects of large-scale implementation of AS of HAI [12], also supplementary to the roadmap.

Healthcare data and standardization

AS relies on reusing routine-care data from EHR. A highly divergent landscape of healthcare software is running in hospitals worldwide, leading to differences in routine-care data. Even if the software is produced by the same company, local adjustments usually result in differences in data structure. In some hospitals, there is no comprehensive EHR, or clinical data are unstructured or distributed across different subsystems. Moreover, these subsystems may apply different coding languages, further limiting their interoperability (section Standardization and interoperability). Implementation of AS requires substantial resources to collect and harmonize data between these sources, in particular if surveillance is intended to be implemented centrally. This is not a one-time investment but a continuous effort, as local data systems are continuously changing. In the following sections, we explain the different types of healthcare data and their structure and how to use them when implementing AS.

Structured and unstructured data

Structured data are characterized by an identifiable format where the content is highly organized in a small and searchable data unit. Common examples of structured data in healthcare are admission/discharge data, ward type, diagnosis codes, procedure codes, some laboratory data like C-reactive protein measurements and vital parameters such as body temperature, all of which are assigned to a unique data type, each with a distinct meaning. This allows the data to be collected, interpreted and validated by algorithms easily and unambiguously.

Unstructured data are recognized as not having a predefined format or organization and are commonly found as free text in EHR, for example doctors' or nurses' notes for patient symptoms. Analysis of unstructured data in AS usually requires transforming it into a structured format, a process that can be complex and resource intensive. Depending on variations in the data content, this can be done by string matching (if uniform spelling) or more advanced text mining techniques such as natural language processing (NLP). NLP methods are becoming increasingly important for making even free-text medical notes machine readable [16]. From an IT perspective, structured data are always preferred over investing in the transformation of unstructured data.

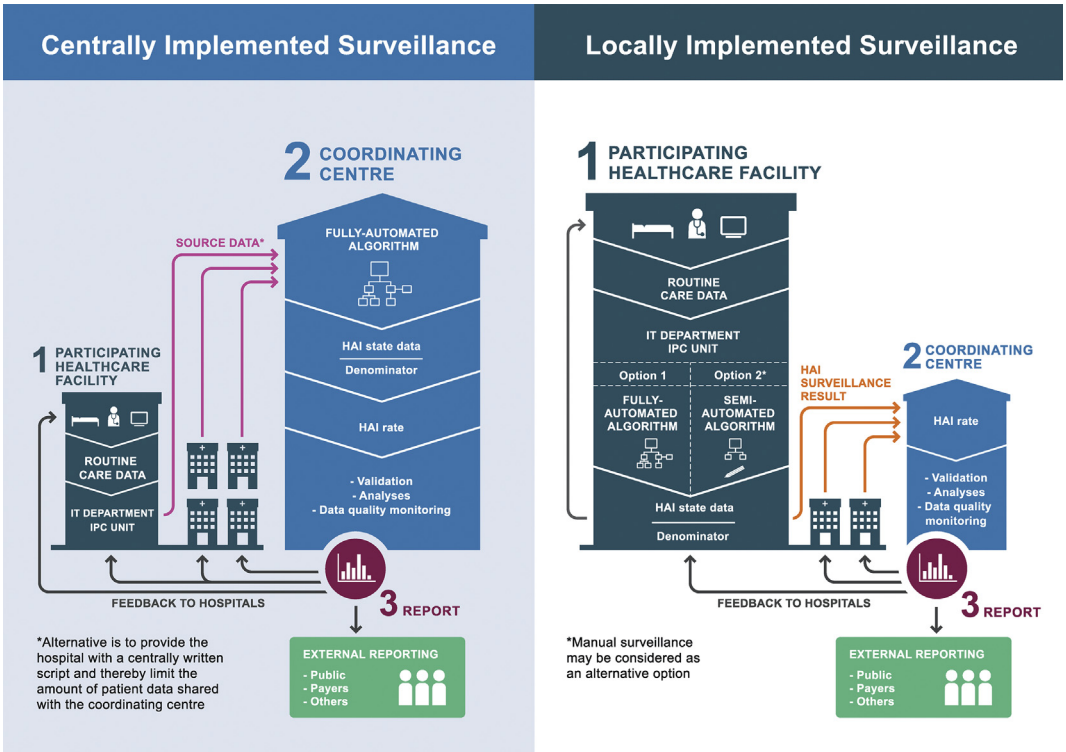

Fig. 1. Schematic representation of approaches to implementing centrally and locally implemented AS. Abbreviations: AS, automated surveillance; HAI, healthcare-associated infection; IPC, infection prevention and control; IT, information technology.

Minimal and health core data set

The required source data for AS comprise a long list of variables and depend on targets of AS and the required algorithms. To facilitate the implementation of AS, a minimal data set (MDS) can be specified [11]. These variables reside in different sources and information systems distributed over the hospital IT infrastructure.

In general, for all fields of application where EHR data are used, the definition of a core data set (CDS) is an important first step for the requirement analysis. The CDS is an agreement by experts defining a set of variables required to apply health algorithms, including algorithms for AS. The CDS is designed as an expandable basic module collection with obligatory specifications, like structure and data type for all included variables. It also reports useful international standards to represent the data (section Standardization and interoperability). Table 1 provides an overview of a health CDS, covering all dimensions of surveillance, not just those pertaining to AS. More detailed information is published by the Medical Informatics Initiative [17].

When building an automated HAI surveillance system, the HAI source data can be organized within a HAI MDS, which is a subset of the CDS including only parameters required for detecting the HAI and the relevant denominator data. For all data in the MDS, the quality of data elements should be assessed. For this purpose, data can be classified into different categories with expanding criteria for accessibility/applicability in algorithms (Table 2). Category 5 applies to data without any digital representation, i.e. health data on paper sheets. In category 1, the data are of ideal quality for AS or other health algorithms. As an example, unstructured microbiology findings data would be in category 4. If these data are structured but the variables do not meet the requirements of international standardization, its category would be 3. The International Classification of Disease (ICD) codes in an Admission Discharge and Transfer (ADT) system matches the criteria of category 1. Table 2 could be used by participating healthcare facilities to check the status quo of digitalization of data needed for the implementation of AS. Ideally, all data needed for implementing AS are in category 1.

Standardization and interoperability

A core requirement for implementing AS is the homogenization of HAI source data format. Transformations of source data to standard format is usually performed by taking small steps at a time, and the choice of standards should be evaluated thoroughly. It is recommended to formulate and communicate a clear transformation strategy, managed by a higher-level organization, like the coordinating centre. These could include e-health initiatives, but a key factor to speed up the overdue process of standardization in digital health is to encourage political legislators to implement mandatory rules that every software company has to fulfil. Valuable information to guide this process can be obtained by looking at similar solutions used by other countries [18] or used at the European level [19].

Syntactic and semantic standardization are the key concepts to practical interoperability. Syntactic interoperability defines the transmission techniques for data, and semantic interoperability is important for unambiguous machine-based understanding and the interpretation of data. Standardization should be achieved at the earliest possible point of data creation. This defines the responsibilities of the companies/departments developing systems in which data are registered to reduce errors and the misinterpretation of data. Ideally, all data which are exchanged between different systems in a hospital setting (involving internal and external computer systems) should be standardized to avoid misinterpretation.

Some existing AS systems use their own defined standards for specific data sets, such as the MiBa (the Danish microbiology database) and the German surveillance network Antibiotika-Resistenz-Surveillance (ARS). The MiBa collects all microbiology findings from all clinical microbiology departments in Denmark, coordinated by the responsible national governmental agency, Statens Serum Institute, and defined by a standard transfer protocol, called MedCom XRPT07 [20]. Because the local data providers use different identifications for pathogens and other data elements, they established a mapping mechanism to convert from local codes to shared codes.

Ideally the source data and/or HAI surveillance result transferred from hospitals to the coordinating centre should be in an international accepted standard format. Data can be transferred using tools like openEHR or ART DECOR as CDA (clinical document architecture) templates or using fast healthcare interoperability resources (FHIR). To standardize the definition of HAI source data and HAI surveillance results, they have to be specified in iteration with stakeholders and domain experts. Content data of the MDS, transportation containers between systems and implemented surveillance algorithms have to be standardized to achieve interoperability among different systems. A specification describing the agreed standardized HAI MDS should be reported publicly. Table 3 shows a selection of important internationally accepted standards for data exchange.

Table 1
Health core data set in a clinical setting

Module	Description	Importance to automatic surveillance
Diagnosis	ICD or similar codes present at admission and discharge	Date, diagnosis codes
Procedures	Procedures in coded format	Surgical procedures: date, procedure code
Laboratory results	All laboratory results, including serology, virology, microbiology	Microbiology findings: date of sample collection, specimen, pathogen, result
Medication	Patient-based medication, including antibiotic treatment	Antibiotic treatment: ATC code, date of prescription, dosage, form of administration, site and route of administration, days of treatment
Person data	To use the module's data in central surveillance networks, these data identify a unique person and can be anonymized	In hospital, patient name or number; in central surveillance networks, pseudonymization or anonymized ID
Demographic data	Birth date, sex, vital status	Year of birth, sex
Case data	Admission, discharge, ward movement, health insurance, address; data are based a patient's visit to the healthcare facility	Date of admission and discharge to specify hospital-acquired infection
Vital signs and score data	Vital sign data (blood pressure, heart rate, temperature, biosigns), scores (SAPS, SOFA)	Infection-relevant signs like fever, date
Structural data	Aggregated data about all aspects of the facility (e.g. bed size, number of employees, cases per year, type of medical departments)	Important for comparison benchmarks to stratify analysis data

Abbreviations: ATC, Anatomical Therapeutic Chemical classification; ICD, International Classification of Disease; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

Table 2
Categories to indicate the suitability of surveillance data in a hospital usable for automated infection surveillance

Surveillance data	Category				
	1	2	3	4	5
Data already exist in a digital subsystem	Yes	Yes	Yes	Yes	No
Data are structured and well defined	Yes	Yes	Yes	No	No
Data are available in most facilities and semantically standardized	Yes	Yes	No	No	No
Data are accessible for surveillance algorithms	Yes	No	No	No	No

To illustrate the use of standards in AS, implementing the HAI algorithms in Arden syntax has the advantage of a unique code interpretation between all surveillance network participants. Its practical use in accessing microbiologic data has been proven [22], but the lack of Arden syntax–standardized queries and data access methods remain a problem [23]. On the other hand, interoperability to openEHR-based data systems has been positively evaluated [24]. Other ways to implement HAI algorithms are common programming languages and the use of application programming interfaces to access the HAI raw data. Some typical barriers in the introduction of standards are illustrated in Box 2.

Box 2
Barriers in introducing e-health standards.

In general, the laboratory information system uses some structure in their databases, but these are not converted to standards like SNOMED CT or LOINC.
If possible, proprietary standards developed by commercial companies should be avoided because licensing can be an issue. For most standards, there are open source solutions available. However, SNOMED CT, which comes with a licence cost, is mostly chosen because of its complexity and integrity.
Because codes are not static but always changing, updates require attention. As an example, the ICD codes are updated annually, and all data sources must also receive these updates, which can be a time-consuming process. Similar problems arise if the hospital decides to update from HL7 v2.x to FHIR and all communicating systems have to be adapted.

Surveillance algorithms

Implementation of algorithms requires several steps in development and validation [25], where close collaboration between software developers, hospital IT staff and domain experts such as clinical physicians and/or IPC professionals is essential. Development can be done by the healthcare facility or the coordinating centre, or it can be outsourced to an external partner. Although universal solutions are possible in theory, all algorithms applied on a new data source require adaptation and validation before being implemented. This can be a complex and time-consuming task.

It is important to not consider the development of algorithms as merely a one-time accomplishment but instead a continuous and iterative process [11]. This requires that, depending on whether surveillance is implemented locally or centrally, the participating healthcare facility and the coordinating centre respectively ensure full access to the code, as well as human resources to adapt the code synchronously with changes in the source data or novel demands from users. Simple algorithms have the advantage of being easy to understand, thereby increasing acceptance among users; most of the time, a simple rule-based classifier provides satisfactory results.

When applied in production, algorithms are usually run in regular time intervals, such as once every 24 hours, once a week or in larger batches. In most circumstances, it is not required that data are being updated in real time. Simpler rule-based classifiers can be written in any language supported in production (e.g. Java, PL/SQL, C++, C#, SAS, R or Python). Specific medical-focused programming languages like Arden syntax could strengthen the exchange of programming code and the common understanding of the algorithms.

Specific considerations for algorithm source data

Most HAI algorithms depend on longitudinal data with time stamps. Data assessment is usually done during a predefined time

Table 3
Selection of important standards in the health information technology sector

Name of standard	Type	Subject
SNOMED CT (Clinical Terms)	Terminology/coding system	Microbiologic content like organisms, specimen and other can be specified with these code systems
LOINC (Logical Observation Identifiers Names and Codes)	Terminology/coding system	See SNOMED CT
UCUM (Unified Code for Units of Measure)	Terminology/coding system	Defines a uniform, international machine-readable reproduction of units of measurement
ICD (International Classification of Diseases)	Terminology/coding system	For coding diagnosis or operational procedures
HL7 v2.x	Data transfer	Messaging standard for syntactical exchanging content to other systems in hospital; based on ASCII text; widely implemented in hospitals
HL7 v3	Data transfer	Supports standardizing of messages (in hospital and external/intersectoral); based on XML
CDA (Clinical Document Architecture)	Tool for structuring	Implements templates of data structures and definitions for exchange purposes
FHIR (Fast Healthcare Interoperability Resources)	Structured tool	A product of the HL7 Foundation and successor to CDA and HL7 v3; easier to handle and more modern (web technics) than v3; still in development
ART DECOR	Structured tool	Open source web-based tool to specify new CDA templates
openEHR	Structured tool	Open standard definition of vendor-independent electronic health record (EHR)
Arden syntax	Programming language	Developed to standardize the representation of medical knowledge for clinical decision support systems; a suitable tool to share and discuss the developed algorithms; used in clinical decision support systems and can fire alerts and other output elements to the clinician or computer systems. Its modules, called MLM (medical logic modules), are procedural units of single medical solutions [21]

window (e.g. symptom data are considered only when linked to a drawn culture or after a surgical procedure) [25]. Some variables could have zero to many measurements during the time window, and it needs to be determined what value to use (e.g. most serious value, latest value, mean value, all values). In EHR, missing data are common, and attempts to address this should take into account the underlying reasons for missing data to avoid biased results [26]. Determining data quality is crucial and needs to involve domain healthcare experts. Even though the EHR may contain features – such as tick boxes for device presence or functionalities to register movements between different wards – healthcare personnel may still not use them in the intended way, despite clear guidelines. Not addressing these issues may severely bias algorithm output.

Technical data validation

Technical validation is essential and needs to be performed before and continuously after implementing an algorithm to identify problems with the quality of source or output data. The validation of a data set is a two-step process. Firstly, the values of all variables of the data set should be checked to assure that it only contains valid values. This validation can be done using different methods. For example, a scheme might be used where the valid values are defined, including definitions of mandatory or optional parameters and valid value ranges. However, if the validations for the values are more complex than simple lists, they can be made with rule-based validation engines. An example of such tools for XML data sets is Schematron [27]. If a record does not fulfil the checks of this step, then it should be considered invalid.

Second is the plausibility validation of the data set. This validation is more complex, as it usually involves more than one record. For example, one plausibility rule for a laboratory finding would be the detection of extreme outlying values or that the sampling date lies within the hospital admission period of the patient episode. If a record does not pass one of the plausibility checks, then it does not automatically mean that the record is invalid. The plausibility checks should have different levels of severity (e.g. Error, Warning or Information) to indicate the urgency to fix the data. This process should use prioritization; a record with an error should be fixed before a record with only a warning.

User interfaces for dissemination

User interfaces are solutions for the end users that allow them to apply the data for their daily work in quality assurance and IPC, and, in the case of public reporting, for citizens to view and understand the data. For quality assurance, visualizations require a comparison over time in order to analyse trends and investigate reasons for changes in HAI rates. Comparing between facilities can be done if data are considered comparable. For IPC departments, it may be preferred that data sets with patient-identifiable data and/or aggregated data are kept in a raw format or at a very detailed aggregation level, so that the data can be analysed according to local needs while allowing for analyzed specific investigations of trends. These detailed investigations will usually require line lists of individual patients.

IT infrastructure design

Hospital subsystems

Many hospitals are overwhelmed by multiple IT systems. Depending on a specific hospital's IT architecture, the source data required are usually distributed among different subsystems. There are several available solutions; the IT architecture can be more or less integrated or connected via communication servers. In this wealth of systems, it is important to realize that similar data may be stored in several systems. In addition, data content is sometimes 'vendor locked' and not accessible, or some data may still be stored on paper. The design of IT systems is influenced by local preferences, possibilities offered by vendors and, if applicable, regulations that enforce specific data standards or protect data privacy.

The main component is usually the hospital information system or EHR, which is often the main system accessed by healthcare professionals. It manages patient care, its administration, ward transfers and billing; it often includes medical data such as procedures, diagnoses, free-text notes and laboratory results. Another important subsystem is the laboratory information system (LIS), which produces and transfers the laboratory results, including microbiologic findings and the pharmacy information system. Some hospitals also use specific order entry systems to allow the clinicians to send orders to subsystems like the LIS. The patient data managing system, mainly implemented at high-dependency units like the intensive care unit, displays vital signs and other patient measurements. Other systems running in a hospital include picture archiving and communications system and radiology information systems.

Routine-care data access and storage

The following section describes approaches to storing and accessing routine-care data from EHRs for use in AS. Data can be shared among multiple participating healthcare facilities, but the description below assumes an approach that keeps all patients' EHR data local in the hospital.

Clinical data repository in hospitals

Routine-care data, consisting of clinical and administration data, are transferred from hospital subsystems into a clinical data repository (CDR), a storage domain including multiple databases and concepts. The CDR includes operational data stores (ODS) for fast writing of real-time data out of clinical data sources and clinical data warehouses (CDW). The CDW are optimized databases for performance of specific fast queries [28]. Different architectures of CDW can be implemented, with the classical CDW including the EHR data and all administrative data, as compared to a research CDW, which has additional functions like patient deidentification for external receivers [29]. Generally, the ODS are source data for the CDW. In addition, data such as microbiologic findings may be transferred from the LIS to the hospital information system and additionally to the CDR. The transmission is controlled by a communication server.

Extract transform load is a method to transfer data. 'Extract' is the process of reading data from different types of sources. 'Transform' is the process of preparing the data for writing and converting the data into the needed structure. 'Load' writes the data into the target database. In the process of transferring the data from the ODS to the CDW via extract transform load, data can be validated and transformed in syntactic and semantic standards if not done before. Data analysis algorithms access the CDW directly, and analysis results can be retransferred to the clinical data documentation in the EHR, if allowed by the vendor.

CDR enhance the transforming process to standardized and interoperable data; this is useful for many data applications, not just for AS. The process of pseudonymization or anonymization for external data access and research should be centralized and can be implemented in a dedicated module as part of the CDR.

Compatibility layer

If the hospital has limited resources for developing the digital infrastructure of a CDR, it can suffice to implement (or buy) a compatibility layer to transform the hospital-specific data required by the MDS specified by the coordinating centre. The source data are stored in a separate database and are accessible for the AS. This layer resembles a minimized CDR, one just to fulfil the requirements of the surveillance network.

Surveillance data centre

The surveillance data centre (SDC) is the computer system located at the coordinating centre; it can be developed for a regional or national surveillance network or an otherwise organized group of hospitals sharing a common surveillance strategy.

The SDC accesses data from participating hospitals, validates data and supports analysis of surveillance data. It contains a communication unit to transfer data to the hospitals and external partners. An example of an SDC is TESSy at the European Center for Disease Prevention and Control (ECDC), to which European countries transmit their HAI data annually [30]. Web services support the secure connection to the external communication partners. A surveillance network SDC includes an ODS and a data warehouse. The hospital data are stored at the ODS and transferred to the data warehouse. Validation is a functional part of this transfer operation. A web-based user interface allows user-defined analysis, reporting and the upload and download of data.

Communication between participating hospitals and coordinating centre

There are many methods to set up the communication infrastructure between healthcare facilities and the SDC to establish AS. This section is divided into parts for centrally and locally implemented AS; example scenarios are discussed in each part. Different scenarios can in general be considered for different surveillance targets. Importantly, different scenarios can be running at the same time in one hospital, especially during the implementation phase. As an example, to begin implementing the AS, the hospital can do the bloodstream infection (BSI) surveillance with the method from one scenario, all while all other conventional surveillance activities are going on as before. The temporary overlapping of surveillance methods can also be taken advantage of for validation purposes of the same target.

Centrally implemented AS

The coordinating centre provides standard definitions of HAI source data, HAI surveillance results, messaging standards and algorithms to identify HAI. Algorithm results (HAI surveillance data) and analysis results of pooled data (HAI rates) are stored in the surveillance network data repository. There are four suggested scenarios to organize centrally implemented AS.

Scenario 1. The HAI source data are transferred from the participating healthcare facilities to the SDC located within the coordinating centre. After the data transfer, the fully automated HAI detection algorithms are applied. Depending on the possibilities regarding data protection, the patient data are anonymized or pseudonymized at the participating healthcare facility level. Analysis reports go back to every hospital; if pseudonyms are created at hospital level, then the link to the patient's name can be reconstructed at the hospital level.

Scenario 2. The HAI source data are not transferred from the participating healthcare facilities to the coordinating centre. Instead, the HAI detection algorithms are running in the SDC, accessing each participating centres' HAI source data in a standardized format via secured application programming interfaces (API). Feedback is organized as in scenario 1.

Scenario 3. Fully automated surveillance algorithms, designed and maintained by the coordinating centre, are running locally on the participating healthcare facilities' CDR. The algorithm outputs (HAI surveillance results) are transferred to the SDC. Feedback is organized as in scenario 1.

Scenario 4. A hybrid method combines semiautomated surveillance with the centralized approach and is divided into different steps, as follows. Firstly, HAI source data are transferred to or accessed by the SDC as in scenario 1 or scenario 2. Secondly, the SDC identifies the suspected HAI and sends this back to the participating healthcare facility. Thirdly, the IPC professionals in the participating healthcare facilities confirm or reject the suspected HAI. Finally, the confirmed HAI are transferred to the SDC with the appropriate state ‘confirmed’.

The healthcare facility's effort to implement the surveillance is lowest in scenario 1. All scenarios require healthcare facilities to collect, standardize and format the HAI source data according to instructions from the coordinating centre. If algorithms and their adjacent IT infrastructure are centrally organized, maintenance is usually low. This means that when code or data structures change in the surveillance algorithms, changes only need to be made within the coordinating centre, and there are no update or distribution efforts.

Locally implemented AS

In a locally implemented AS, a greater IT responsibility lies within the participating healthcare facility. There are two suggested scenarios to organize locally implemented AS.

Scenario 1. Semiautomated surveillance algorithms are implemented in each participating healthcare facility, creating ‘suspected HAI events’. The IPC professionals rate these events, and the HAI surveillance results are transferred to the SDC at the coordinating centre.

Scenario 2. Locally implemented fully automated algorithms classify HAI and save the HAI surveillance results into local databases. Hospital-specific indicators can optionally be included in the surveillance and local analysis process. HAI surveillance results are transferred to the SDC. The concept of the algorithms and the definition of the core HAI surveillance result data are adapted from the coordinating centre.

Secured data transfer

Secure transfer of surveillance data is a priority in a surveillance network. Whenever possible, data should be pseudonymized or anonymized, and only the required data for the purpose of surveillance should be shared. Data transfers need to take place using secure channels, especially when data are transferred from participating healthcare facilities to the coordinating centre via public internet. Within the healthcare facilities, the data flow needs to be secured with the established local security standards. This requires that all connections be recorded, and unauthorized access trials needs be communicated from the system to the responsible administrators.

Data transfer from participating healthcare facilities to coordinating centre

Participating healthcare facilities and the coordinating centre are usually not in the same geographic location and do not share the same private or internal network. The easiest solution for data transfer is to use the internet to communicate between both sites. One of the most commonly used protocols is hypertext transfer protocol secure (HTTPS), particularly in fully AS systems, where API could be put in place or via an upload form. Alternatives to transfer data from the participating healthcare facilities to the coordinating centre include secure file transfer protocol (SFTP) and secure shell (SSH); data can also be encrypted if necessary. Another option is to use a dedicated network or a virtual private network (VPN), which requires customization. In some countries, like Sweden, a dedicated network has been put in place like the Sjunet (Swedish Health Care Network) comprising an infrastructure for communication between hospitals, primary care centres and home care [31].

Data transfer from coordinating centre to participation centre

The coordinating centre should provide feedback to the participating centre on the infection rates and other information, like validation and analysis results. Web portals with HTTP and authentication mechanisms are required so that every participating centre can access its data individually. When the aggregated data become public, no authentication mechanism is required, and (aggregated and pooled) infection rates are available on the internet via a public page or a report in PDF form.

Practical example: a minimal viable product approach to set up locally implemented AS

The minimal viable product approach means development of AS with just enough features to satisfy the basic needs of HAI surveillance and to facilitate feedback from users before proceeding further in development. The methodology is usually an efficient first step when AS is being implemented from scratch or when new surveillance targets are being introduced. To illustrate this process in the context of automated HAI surveillance, we have used surveillance of hospital-onset bacteraemia (Box 1) in a fictitious locally implemented AS system, here called automated in-house surveillance system (AISS), as an example. The schematic structure of the fictive AS system is illustrated in Fig. 2.

This AISS example has many advantages. Firstly, it would work as a stand-alone software, so it would not disturb the existing hospital IT infrastructure. Secondly, it can be developed by specialized medical informatics or purchased as third-party software (Box 3). The hospital IT employees have to provide the read access for the AISS to the hospital subsystems and the server infrastructure running AISS.

After receiving feedback from users, and if more data from the subsystems are available, the system can be elaborated in further stages on various dimensions, as outlined in Box 4.

Box 3
Commercial versus open source software.

All scenarios can also be realized with third-party software, either locally integrated into the user interface of the hospital information system (HIS) or as a stand-alone system. This software should obtain the source data for automated surveillance (AS) directly from the HIS infrastructure. Also, it has an interface to the surveillance data centre to exchange healthcare-associated infection (HAI) result data and receive analysis data. A preferred approach is to release this kind of software under a free or open source licence to ensure data independence and flexibility. With upgrades, new functions and licence fees, commercial software can be expensive in the long term. The open source approach also has the advantage of facilitating centres in low- and middle-income settings to start with surveillance activities. Software developed by several people, ideally collaboratively, ensures its continued existence and maintenance. Applications created by a single person are usually not durable in the long run and should be avoided. The basic principle for in-house development is to transfer the knowledge to as many people as possible in order to minimize the risk of losing all knowledge. But even if you choose open source software, you are not safe from later commercialization or termination of the product; for example, the collaboration tool DekiWiki (later called MindTouch Core), which was released by a company as open source software, discontinued development in 2013 [32].

Box 4
Dimensions of system elaboration.

Expand the MDS to collect more source data to improve algorithms or collect more risk-factor data.
Target additional HAI, e.g. surgical site infections, perhaps needing more source data (e.g. procedures, antibiotic use).
Target specific patient populations (e.g. ward level, devices).
Use more complex algorithms (e.g. with more complex methods).
Build in a possibility for semiautomated surveillance.
Ensure the participation of multiple participating healthcare facilities.
Automate the reporting to and from the coordinating centre to the participating centre.

Maintenance and quality control

The types of surveillance systems described above require continuous maintenance and quality assurance in order to ensure reliable output. Maintenance requires effective governance, and guidance can be found in the information technology infrastructure library (ITIL), which is a detailed description of practices in IT service management. Two important concepts in ITIL are incident management and change management. Incident management is needed to handle problems with the system. This includes a monitoring system to identify problems and their nature as well as a service desk/hotline with a ticket system and agreements on second- and third-level support. Change management is a process used to collect and prioritize suggestions and wishes for improvements and implement them in a structured way.

In addition, algorithms have to be evaluated regularly and adapted when necessary. The source data used for the algorithms are dynamic both in data registration and in their implementation. As an example, a department of clinical microbiology may change their codes for blood cultures in the microbiology database. As a result, the algorithm for BSI no longer picks up these blood culture results, resulting in an apparent drop in the incidence of BSI in that participating facility. This issue may be minimized, but not eliminated, by using standards such as SNOMED CT or LOINC. A more

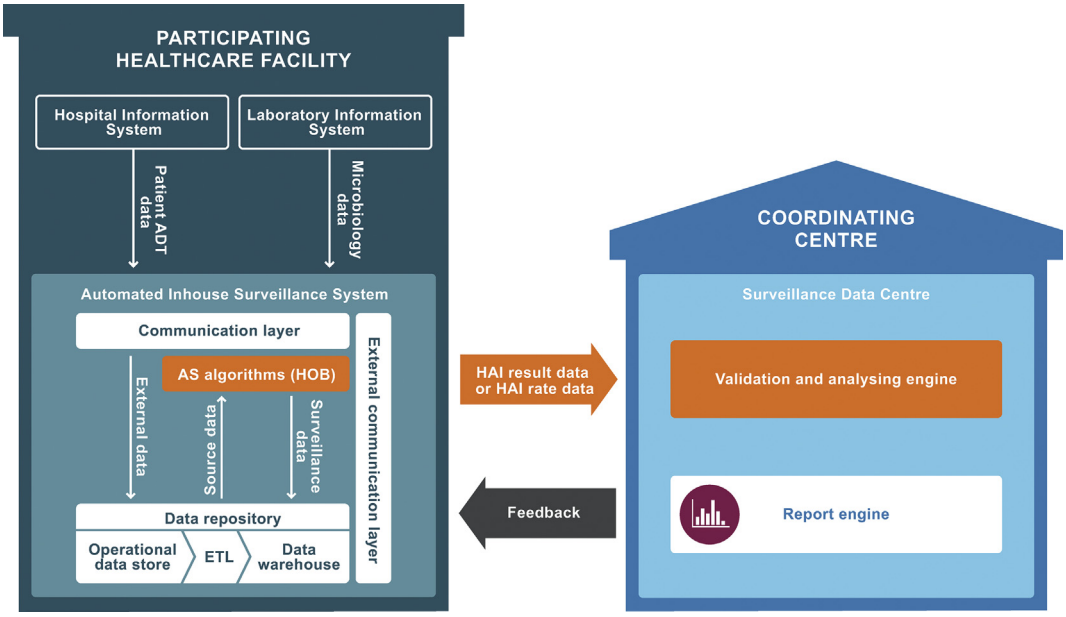

Fig. 2. A minimal viable product approach to set up locally implemented automated surveillance. The AISS reads patient ADT from the HIS and receives microbiology findings from the hospital's LIS. These data are stored into the operational data store. ETL processes validate, standardize and transform these data to a data warehouse. The HOB detection algorithm (Box 1) is running locally in the AISS and performs the classification on the basis of the data transferred from HIS and LIS. The code template for the algorithm is provided by the coordinating centre, which also sends update notes to all participants, although in the local facility the algorithm can be implemented using any programming language. To ensure the quality of the implemented code, the coordinating centre also offers input and output test data for the local algorithms. The HAI surveillance result data, or alternatively just the aggregated HAI rate, are transferred to the SDC. Which data are required to send depends on the strategy chosen by the coordinating centre. The SDC validates and analyses the data and sends feedback to the participant's AISS. Reports are not transferred from the SDC to the AISS because they are accessible via web interface directly from the SDC. Abbreviations: ADT, admission, discharge and transfer system; AS, automated surveillance; ETL, extract transform load; HIS, hospital information system; HOB, hospital-onset bacteraemia; LIS, laboratory information system; MIBI, microbiology; SDC, surveillance data centre.

rigorous example of a change in registration would be a complete revision of the data model of the data source, requiring new data importers and a revision and test of all algorithms. Changes in implemented surveillance systems or electronic healthcare registration systems may involve moving a database to another server, requiring new setup, new user access, updating scripts and updating agreements.

These changes are for the most part not within the domain of the team maintaining the AS system. In order to identify changes and respond to them in a timely manner, there are a few methods that can be used in parallel. In the first place, an active dialogue with all stakeholders is required in order to be informed about changes. If possible, it is helpful to be involved in the design of the changes to make sure that they can be incorporated in the surveillance systems. This is, however, not a watertight method, and in practice, it can be expected that the team will be informed much later or not at all. Therefore, secondly, an automated monitoring of trends on specific source data at the hospital and department levels will be useful to receive notice of unexpected changes. Thirdly, the surveillance data can be compared to reference data, such as data generated by manual evaluation of patient records, every year or every few years. Many health registries are primarily maintained by IT specialists, but the complexity and level of detail in these AS systems require both IT specialists and experts in IPC, microbiology and epidemiology to be involved in maintenance. In addition, it is important to hire experts for the long term. These systems are complex and cannot be run by a team with high turnover and its attendant loss of experience and expertise.

Backups and version management, for example monthly, are helpful when specific trends need to be investigated. These backups may include extracts from the data sources and the coding of the algorithms and/or output data. The latter can also be reproduced if the first two are stored. Issues to consider are the legal basis for storing data and the need for storage space.

Future directions

By creating proof-of-concept systems, preferably developed as open source stand-alone systems, large-scale AS networks can be used as best practices for healthcare facilities and can contribute to the development of new systems. Analysing the feedback of the people involved should specifically focus on IT-based experiences for system-to-system communication and infrastructure, as well as the collaboration between the IT department and IPC professionals. To increase standardization, companies selling LIS, specialized patient data managing system for intensive care units and similar medical subsystems should offer default APIs that are based on international interoperability standards. This allows for the development of more advanced surveillance algorithms with automatic processing of structured data. In addition, AS needs to develop in parallel with data legislations. Implementing a general framework, which allows AS within the existing General Data Protection Regulation, is crucial to allow for surveillance networks to collect and share desired HAI surveillance data regionally, nationally and on the European level. Finally, the initiative and leadership of automated HAI surveillance should be maintained by the IPC community and medical informatics experts. This can be ensured by establishing a commercially independent organization with the objective of compiling information on worldwide AS activities, implementing guidelines and pushing the development of AS further.

Transparency declaration

This network has been supported under the 7th transnational call within the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR), Network Call on Surveillance (2018) and was thereby funded by ZonMw (grant 549007001). This project also received support from the COMBACTE MAGNET EPI-Net project funded by the Innovative Medicines Initiative Joint Undertaking under grant agreement 115523 | 115620 | 115737 | 777362, resources of which are composed of financial contribution from the European Union Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution. J.K.V. was supported by grants from Region Stockholm and Vinnova. The other authors report no conflicts of interest relevant to this article.

Acknowledgements

We would like to thank E. Abbevik for her help in designing the graphical work. We are thankful for the valuable input of our external reviewers: K. Woeltje (**USA**), A. Harris (**USA**), A. Hajdu (**Hungary**), B. Babarczy (**Hungary**), T. Karki (**Sweden**) and C. Schweizer (**Germany**).

Members of the PRAISE network are: Olov Aspevall, Unit for Surveillance and Coordination, Public Health Agency of Sweden, Solna, **Sweden**; Pascal Astagneau, Centre for Prevention of Healthcare-Associated Infections, Assistance Publique – Hôpitaux de Paris & Faculty of Medicine, Sorbonne University, Paris, **France**; Marc J. M. Bonten, Department of Medical Microbiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, **The Netherlands**; Elena Carrara, Infectious Diseases Section, Department of Diagnostics and Public Health, University of Verona, Verona, **Italy**; Aina Gomila-Grange, Infectious Diseases Unit, Bellvitge Biomedical Research Institute (IDIBELL), Bellvitge University Hospital and Infectious Diseases Unit, Consorci Corporació Sanitària Parc Taulí, Barcelona, **Spain**; Sabine C. de Greeff, Centre for Infectious Disease Epidemiology and Surveillance National Institute for Public Health and the Environment (RIVM), Bilthoven, **The Netherlands**; Wendy Harrison, Healthcare Associated Infections, Antimicrobial Resistance and Prescribing Programme (HARP), Public Health Wales, Bangor, Wales, **UK**; Hilary Humphreys, Department of Clinical Microbiology, The Royal College of Surgeons in Ireland, Department of Microbiology, Beaumont Hospital, Dublin, **Ireland**; Anders Johansson, Clinic for Infectious Diseases, Umea University, Umea, **Sweden**; Mayke B.G. Koek, Department of Epidemiology and Surveillance, Centre for Infectious Disease Control, National Institute for Public Health and Environment (RIVM), Bilthoven, **The Netherlands**; Alain Lepape, Clinical Research Unit, Department of Intensive Care, Centre Hospitalier Universitaire Lyon Sud, Pierre-Bénite, **France**; Jean-Christophe Lucet, Infection Control Unit, Hôpital Bichat-Claude Bernard Assistance Publique – Hôpitaux de Paris, Paris, **France**; Siddharth Mookerjee, Infection Prevention and Control department, Imperial College Healthcare NHS Trust, London, England, **UK**; Pontus Naucler, Department of Medicine Solna, Division of Infectious Diseases, Karolinska Institutet and Department of Infectious Diseases, Karolinska University Hospital, Stockholm, **Sweden**; Zaira R. Palacios-Baena, Unit of Infectious Diseases, Clinical Microbiology and Preventive Medicine, Hospital Universitario Virgen Macarena, Institute of Biomedicine of Seville (IBIS), Seville, **Spain**; Elisabeth Presterl, Department of Infection Control and Hospital Epidemiology, Medical University of Vienna, Vienna, **Austria**; Miquel Pujol, Bellvitge University Hospital, Barcelona, **Spain**; Jacqui Reilly, Safeguarding Health Through Infection Prevention Research Group, Institute for Applied Health Research, Glasgow Caledonian University, Glasgow, Scotland, **UK**; Christopher Roberts, Healthcare Associated Infections, Antimicrobial Resistance and Prescribing Programme (HARP), Public Health Wales, Bangor, Wales, **UK**; Evelina Tacconelli, Infectious Diseases, Research Clinical Unit, DZIF Center, University Hospital Tübingen, Tübingen, Germany and Infectious Diseases Section, Department of Diagnostics and Public Health, University of Verona, Verona, **Italy**; Thomas Tängdén, Department of Medical Sciences, Uppsala University, Uppsala, **Sweden**;

References

[1] Umscheid CA, Mitchell MD, Doshi JA, Agarwal R, Williams K, Brennan PJ. Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. Infect Control Hosp Epidemiol 2011;32:101–14.
[2] Cassini A, Plachouras D, Eckmanns T, Abu Sin M, Blank HP, Ducomble T, et al. Burden of six healthcare-associated infections on European population health: estimating incidence-based disability-adjusted life years through a population prevalence-based modelling study. PLoS Med 2016;13:e1002150.
[3] Suetens C, Latour K, Kärki T, Ricchizzi E, Kinross P, Moro ML, et al. Prevalence of healthcare-associated infections, estimated incidence and composite antimicrobial resistance index in acute care hospitals and long-term care facilities: results from two European point prevalence surveys, 2016 to 2017. Euro Surveill 2018;23:1800516. Erratum in: Euro Surveill 2018;23.
[4] Storr J, Twyman A, Zingg W, Damani N, Kilpatrick C, Reilly J, et al. Core components for effective infection prevention and control programmes: new WHO evidence-based recommendations. Antimicrob Resist Infect Control 2017;6:6.
[5] Gubbels S, Nielsen J, Voldstedlund M, Kristensen B, Schønheyder HC, Ellermann-Eriksen S, et al. National automated surveillance of hospital-acquired bacteremia in Denmark using a computer algorithm. Infect Control Hosp Epidemiol 2017;38:559–66.
[6] Abbas M, de Kraker MEA, Aghayev E, Astagneau P, Aupee M, Behnke M, et al. Impact of participation in a surgical site infection surveillance network: results from a large international cohort study. J Hosp Infect 2019;102:267–76.
[7] Russo PL, Shaban RZ, Macbeth D, Carter A, Mitchell BG. Impact of electronic healthcare-associated infection surveillance software on infection prevention resources: a systematic review of the literature. J Hosp Infect 2018;99:1–7.
[8] Schroder C, Behnke M, Gastmeier P, Schwab F, Geffers C. Case vignettes to evaluate the accuracy of identifying healthcare-associated infections by surveillance persons. J Hosp Infect 2015;90:322–6.
[9] van Mourik MSM, Perencevich EN, Gastmeier P, Bonten MJM. Designing surveillance of healthcare-associated infections in the era of automation and reporting mandates. Clin Infect Dis 2018;66:970–6.
[10] Streefkerk HRA, Verkooijen RP, Bramer WM, Verbrugh HA. Electronically assisted surveillance systems of healthcare-associated infections: a systematic review. Euro Surveill 2020;25:1900321.
[11] van Mourik MSM, van Rooden SM, Abbas M, Aspevall O, Astagneau P, Bonten MJM, et al. PRAISE: providing a roadmap for automated infection surveillance in Europe. Clin Microbiol Infect 2021;27:S3–19. https://doi.org/10.1016/j.cmi.2021.02.028.
[12] van Rooden SM, Aspevall O, Carrara E, Gubbels S, Johansson A, Lucet JC, et al. Governance aspects of large-scale implementation of automated surveillance of healthcare-associated infections. Clin Microbiol Infect 2021;27:S20–8. https://doi.org/10.1016/j.cmi.2021.02.026.
[13] Lin MY, Woeltje KF, Khan YM, Hota B, Doherty JA, Borlawsky TB, et al. Multicenter evaluation of computer automated versus traditional surveillance of hospital-acquired bloodstream infections. Infect Control Hosp Epidemiol 2014;35:1483–90.
[14] Rock C, Thom KA, Harris AD, Li S, Morgan D, Milstone AM, et al. A multicenter longitudinal study of hospital-onset bacteremia: time for a new quality outcome measure? Infect Control Hosp Epidemiol 2016;37:143–8.
[15] Sips ME, Bonten MJM, van Mourik MSM. Semiautomated surveillance of deep surgical site infections after primary total hip or knee arthroplasty. Infect Control Hosp Epidemiol 2017;38:732–5.
[16] Werff SDV, Thiman E, Tanushi H, Valik JK, HenrikssonA, Alam MU, et al. The accuracy of fully-automated algorithms for surveillance of healthcare-associated urinary tract infections in hospitalized patients. J Hosp Infect 2021:139–47.
[17] Medical Informatics Initiative. MII core data set. MII; 2020. https://www.medizininformatik-initiative.de/sites/default/files/inline-files/MII_04_Core_Data_Set_1-0.pdf. [Accessed 1 October 2020].
[18] NSW Government. eHealth strategy for NSW Health 2016–2026. NSW Government; 2015. https://www.health.nsw.gov.au/ehealth/documents/ehealth-strategy-for-nsw-health-2016-2026.pdf. [Accessed 1 October 2020].
[19] European Commission. European eHealth interoperability roadmap. December 2010. http://www.ehgi.eu/Download/European%20eHealth%20Interoperability%20Roadmap%20%5BCALLIOPE%20-%20published%20by%20DG%20INFSO%5D.pdf. [Accessed 1 October 2020].
[20] Voldstedlund M, Haarh M, Mølbak K, MiBa Board of Representatives. The Danish microbiology database (MiBa) 2010 to 2013. Euro Surveill 2014;19:20667.
[21] Pryor TA, Hripcsak G. The Arden syntax for medical logic modules. Int J Clin Monit Comput 1993;10:215–24.

[22] de Bruin JS, Koller W, Zeckl J, Blacky A, Rappelsberger A, Adlassnig KP. Arden syntax MLM building blocks for microbiological concepts and their application in infection surveillance. Stud Health Technol Inform 2017;236:16–23.
[23] Arkad K, Gao XM, Ahlfeldt H. Query-handling in MLM-based decision support systems. Med Inform (Lond) 1995;20:229–40.
[24] Spineth M, Rappelsberger A, Adlassnig KP. Achieving interoperability between Arden-syntax–based clinical decision support and openEHR-based data systems. Stud Health Technol Inform 2018;248:338–44.
[25] Valik JK, Ward L, Tanushi H, Mullersdorf K, Ternhag A, Aufwerber E, et al. Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: observational study using electronic health records data. BMJ Qual Saf 2020;29:735–45.
[26] Hu Z, Melton GB, Arsoniadis EG, Wang Y, Kwaan MR, Simon GJ. Strategies for handling missing clinical data for automated surgical site infection detection from the electronic health record. J Biomed Inform 2017;68:112–20.
[27] van der Vlist E Schematron, Sebastopol CA. O'Reilly. 2007.
[28] Nadkarni P. Clinical data repositories: warehouses, registries, and the use of standards. In: Clinical research computing: a practitioner's handbook. Amsterdam: Elsevier; 2016. p. 173–85.
[29] Shin SY, Kim WS, Lee JH. Characteristics desired in clinical data warehouse for biomedical research. Healthc Inform Res 2014;20:109–16.
[30] Ammon A, Faensen D. Surveillance of infectious diseases at the EU level. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2009;52:176–82.
[31] Inera.se. Sjunet – communications network. 2020. https://www.inera.se/Sjunet. [Accessed 1 October 2020].
[32] MindTouch. Wikipedia. 2020. https://en.wikipedia.org/. [Accessed 22 February 2021].

Glossary

ADT	:	Admission, discharge and transfer system
Arden syntax	:	Programming language for medical purposes
ARS	:	Antibiotika-Resistenz-Surveillance
ART DECOR	:	Open-source web-based tool to specify new CDA templates; the Austrian nationwide electronic health record (ELGA) is specified with ART DECOR
AS	:	Automated surveillance
ATC	:	Anatomical Therapeutic Chemical (ATC) classification system
BCoDE	:	Burden of communicable diseases in Europe
BSI	:	Bloodstream infection
C#	:	Programming language
C++	:	Programming language
CDA	:	Clinical document architecture
CDR	:	Clinical data repository
CDS	:	Core data set
CDW	:	Clinical data warehouse
Change management	:	Technical and organizational process to change IT processes
CVC	:	Central vascular catheter
Data warehouse	:	Optimized databases for performance, reporting and specific fast queries
Docker	:	Tool which supports virtualization of applications on an operating system level
ECDC	:	European Center for Disease Prevention and Control
EHR	:	Electronic health record
ELGA	:	Austrian nationwide electronic health record
ETL	:	Extract transform load, a data transfer method
FHIR	:	Fast healthcare interoperability resources
HAI	:	Healthcare-associated infection
HIS	:	Hospital information system
HL7	:	Health Level 7, a group of international standards for data exchange of healthcare organizations
HOB	:	Hospital-onset bacteraemia (bloodstream infection)
HTTPS	:	Hypertext transfer protocol secure, used for secure communication on the internet
ICD	:	International Classification of Disease
ICU	:	Intensive care unit
Incident management	:	Technical and organizational process to solve IT incidents
IPC	:	Infection prevention and control
IT	:	Information technology
ITIL	:	Information technology infrastructure library, a detailed description of practices in IT service management
Java	:	Programming language
JPIAMR	:	Joint Programming Initiative Antimicrobial Resistance
LIS	:	Laboratory information system
LOINC	:	Logical observation identifiers names and codes
MDS	:	Minimal data set
MiBa	:	Danish microbiology database
MVP	:	Minimal viable product
NLP	:	Natural language processing
ODS	:	Operational data stores
OpenEHR	:	Open standard definition of vendor-independent electronic health record, a suite of concepts and tools
PACS	:	Picture archiving and communication system
PDMS	:	Patient data management system
PHIS	:	Pharmacy information system
PL/SQL	:	Programming language
PPS	:	Point prevalence survey
PRAISE	:	Providing a Roadmap for Automated Infection Surveillance in Europe
Python	:	Programming language
R	:	Programming language
RIS	:	Radiology information system
SAPS	:	Simplified Acute Physiology Score
SAS	:	Suite of analytic software created by the SAS Institute
Schematron	:	Rule-based validation language
SDC	:	Surveillance data centre
SFTP	:	Secure file transfer programme
Sjunet	:	Swedish healthcare network
SNOMED CT	:	Collection of clinical terms which are systematically structured and suitable for machine processing
SOFA	:	Sequential Organ Failure Assessment
SSH	:	Secure shell, a secure network protocol
SSI	:	Surgical site infection
TESSy	:	The European Surveillance System
UCUM	:	Unified code for units of measure
VPN	:	Virtual private network
XML	:	Extensible markup language, a computer language used to define rules for structuring documents