

REVIEW

Open Access



Federated systems for automated infection surveillance: a perspective

Stephanie M. van Rooden^{1*}, Suzanne D. van der Werff^{2,3}, Maaïke S. M. van Mourik⁴,
Frederikke Lomholt⁵, Karina Lauenborg Møller⁵, Sarah Valk¹, Carolina dos Santos Ribeiro⁶,
Albert Wong⁷, Saskia Haitjema⁸, Michael Behnke^{9,10} and Eugenia Rinaldi¹¹

Abstract

Automation of surveillance of infectious diseases—where algorithms are applied to routine care data to replace manual decisions—likely reduces workload and improves quality of surveillance. However, various barriers limit large-scale implementation of automated surveillance (AS). Current implementation strategies for AS in surveillance networks include central implementation (i.e. collecting all data centrally, and central algorithm application for case ascertainment) or local implementation (i.e. local algorithm application and sharing surveillance results with the network coordinating center). In this perspective, we explore whether current challenges can be solved by federated AS. In federated AS, scripts for analyses are developed centrally and applied locally. We focus on the potential of federated AS in the context of healthcare associated infections (AS-HAI) and of severe acute respiratory illness (AS-SARI). AS-HAI and AS-SARI have common and specific requirements, but both would benefit from decreased local surveillance burden, alignment of AS and increased central and local oversight, and improved access to data while preserving privacy. Federated AS combines some benefits of a centrally implemented system, such as standardization and alignment of an easily scalable methodology, with some of the benefits of a locally implemented system including (near) real-time access to data and flexibility in algorithms, meeting different information needs and improving sustainability, and allowance of a broader range of clinically relevant case-definitions. From a global perspective, it can promote the development of automated surveillance where it is not currently possible and foster international collaboration. The necessary transformation of source data likely will place a significant burden on healthcare facilities. However, this may be outweighed by the potential benefits: improved comparability of surveillance results, flexibility and reuse of data for multiple purposes. Governance and stakeholder agreement to address accuracy, accountability, transparency, digital literacy, and data protection, warrants clear attention to create acceptance of the methodology. In conclusion, federated automated surveillance seems a potential solution for current barriers of large-scale implementation of AS-HAI and AS-SARI. Prerequisites for successful implementation include validation of results and evaluation requirements of network participants to govern understanding and acceptance of the methodology.

Keywords Automated surveillance, Healthcare associated infections, Severe acute respiratory infections, Federated systems, Interoperability, Standards, Implementation, Ethical Legal Societal Implications, Governance

*Correspondence:

Stephanie M. van Rooden
stephanie.van.rooden@rivm.nl

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Surveillance is defined as the continuous, systematic collection, analysis and interpretation of health-related data, [1, 2] and surveillance programs for infectious diseases are essential for infection control and pandemic preparedness. For example, they are considered by the World Health Organization (WHO) as one of eight essential components for effective prevention of healthcare associated infections (HAI) [3, 4]. HAI are infections occurring during the process of medical care, with an estimated prevalence of 6.5% of patients admitted in European acute care healthcare facilities [5]. By providing feedback of surveillance results, national and healthcare facility-level trends in incidence rates of HAI and its determinants can be evaluated and benchmarked; thereby HAI surveillance serves as a guidance for infection prevention measures and a method to evaluate the effect of these measures. Currently, automated HAI surveillance is mainly performed in high income countries, although varying in extent between surveillance targets. Limitations in infrastructure, expertise, time and investments often impede structural surveillance in lower income countries, [6, 7] although there are initiatives in these countries to systematically collect HAI surveillance data for a longer period, such as from an international consortium through an online platform [8].

Surveillance of community-onset diseases, such as influenza-like illness (ILI) or community-acquired pneumonia is also important to recognize epidemics and to take responsive measures by public health or healthcare facilities. The COVID-19 pandemic has stressed the importance of continuous high-quality surveillance of infections [9, 10]. More than a decade earlier, the Influenza A(H1N1) pandemic already elucidated the need for high-quality, timely, and long-term surveillance of Severe Acute Respiratory Infection (SARI), a need that was confirmed during [11] the influenza epidemic of 2015/2016 [10]. Influenza surveillance integrated through influenza-like illness/severe acute respiratory infection (ILI/SARI) surveillance platforms, automated and non-automated, is limited globally due to limitations in sustained resources, coordination and commitment, and timely, transparent sharing of good quality information [12].

Traditionally, surveillance is performed in healthcare facilities by manual review of patient records to assess whether the patient meets a standardized case-definition for an infection. The outcome of this surveillance is subsequently shared with regional or national coordinating centers according to fixed specifications. This classical method is error prone [13] and resource intensive [14], limiting timeliness and long-term surveillance performance. Over the last 30 years, a progressive digitalization of information in healthcare facilities has taken place. The widespread

adoption of electronic health records (EHRs) offers the opportunity to reuse routine care data for automation of surveillance. This way, the selection of the surveillance population of interest and the deployment of algorithms to identify (probable) cases can be automated [15, 16]. This automated surveillance (AS) may lead to reduction of workload, while increasing accuracy and completeness. Automation involves increased timeliness and improved standardization, thereby generating an improved quality and utility of surveillance results [15, 17, 18]. Not surprisingly, many algorithms for AS have been developed in the past decades. But these initiatives were mainly confined to research settings, with implementation limited to the healthcare facility level. There are only a few examples of national surveillance networks with implemented AS [19, 20].

When moving AS from the research setting to large-scale implementation in surveillance networks, several important choices need to be made, [21] to identify the most suitable and successful strategy [22]. As will be described, current implementation strategies carry their own set of limitations and challenges, including privacy issues, (data) technical limitations, absence of knowledge and resources needed to deploy AS systems in individual healthcare facilities with the local implementation approach. These limitations will likely even have larger impact when implementing more complex algorithms for case ascertainment, and (rapid) upscaling to other (surveillance) targets, for enhancement of pandemic preparedness.

A new approach may solve the limitations of existing implementation strategies; with federated analyses, scripts for analyses are developed centrally in a coordinating center but applied locally in healthcare facilities. This solution has been tested in studies and consortia with various purposes and may also be a solution for automated infection surveillance.

In this perspective, the potential of a federated system for AS of infectious diseases will be explored in the context of HAI and SARI surveillance. These surveillance targets will serve as an example of AS where clinical data is reused to explore potential applications. We will first give a concise definition of AS and a description of strategies and needs for large-scale implementation, then explain the concept of a federated system in more detail. We will discuss prerequisites for implementation and consider whether federated systems can be a suitable implementation strategy for AS.

Automated surveillance and current strategies for implementation

AS is defined as any form of surveillance where (parts of) the manual assessment is replaced by an automated process, as described in detail by the PRAISE roadmap. [22] In AS, specific algorithms are applied to so-called

source data, a defined subset of data collected during the routine process of care and documented in EHRs. AS systems can be semi-automated or fully-automated. Semi-AS systems rely on human interpretation of results to make the final ascertainment of an infection according to the case-definition for selected cases, whereas fully-automated systems do not require human intervention and ascertainment is completely performed by algorithms.

In HAI surveillance, a specific surveillance population—for example a selection of patients that underwent a specific surgical procedure, or had a medical device inserted—is evaluated to assess whether patients meet the case-definition of a specific type of HAI. In the context of automated HAI surveillance (AS-HAI), semi-automated surveillance (semi-AS) uses algorithms to identify patients with a high probability of having developed the HAI of interest. Only these patients undergo manual medical record review. Most patients are classified as having a low probability of having developed HAI and are assumed free of HAI. Semi-AS requires little to no adaptation of traditional case-definitions as the manual confirmation step allows for incorporation of clinical judgement. In fully-automated surveillance (fully-AS), case-definitions are often adapted to exclude clinical signs and symptoms or rely on proxy-indicators such as diagnosis codes to enable automated decisions [15, 19, 23]. Common types of source data used in AS-HAI are microbiology results, antibiotic prescriptions, admission

and discharge dates, days of central line insertion and procedure codes.

In SARI surveillance, all patients admitted to a health-care facility are assessed for the presence of a clinical syndrome meeting the case-definition of SARI, with national population or a catchment area of a healthcare facility considered the numerator. Semi-AS of SARI relies on manual identification of cases fulfilling the WHO SARI case definition [24] or a similar clinical case-definition in a selection of patients who were automatically flagged based on diagnosis codes [25]. Fully-AS SARI surveillance systems rely on a proxy case-definition, often based on diagnosis codes given at admission or discharge. Sometimes, although case ascertainment is done fully-automated, manual steps are still required for enrichment of additional detailed data for further analyses (e.g. patient characteristics, symptoms or pathogens). For AS-SARI, examples of source data are diagnosis codes from admission or discharge, admission and discharge date and microbiological results. Table 1 summarizes the most important aims and requirements for HAI and SARI surveillance.

Implementation strategies for large-scale automated surveillance

Surveillance can be performed in a local healthcare setting or within a surveillance network, depending on the use case of surveillance and on what level interventions will be implemented to prevent infections (e.g.

Table 1 Aims and requirements for HAI and SARI surveillance

	HAI	SARI
Aims of surveillance	Provide insight into infectious complication of medical care to initiate or evaluate preventive interventions; timely detection of increased rates of HAI, possibly related to breaches in infection control	Monitor epidemiologic trends and characteristics and impact of SARI on healthcare facility utilization and mortality Early detection of outbreaks of known or unknown pathogens Evaluate public health interventions such as vaccination and monitor vaccine effectiveness Identify risk factors for SARI
Numerator, case definition	Standardized case-definitions, usually including microbiological testing outcomes with clinical signs and symptoms	Syndromic case-definitions either fulfilling the WHO case definition for SARI or a proxy case-definition. Ideally complemented with causative pathogens
Denominator	Patient days at risk, device-days at risk or number of procedures	General population or catchment population of participating centres
Large-scale uniformity of case definition and case ascertainment	Important, for assessing local performance against comparators in time or place	Important for comparisons between countries or catchment populations and comparisons in time
Reliable over time	Very important, to detect changes in relatively rare outcomes	Important to detect small changes due to emerging pathogens
Timeliness	Less important, retrospective by design	Very important for the aim of early detection to guide control measures for epidemics
Risk adjustment	Important, for assessing local performance against comparators in time or place	Important for emerging pathogens to identify risk groups and to adjust for differences in catchment populations on a national level

local quality improvement or evaluation of public health interventions). If focusing on implementation of AS on a larger scale, within a regional or national surveillance network, generally, two strategies have been described. AS can either be implemented centrally, where a coordinating center collects all unprocessed data from participating centers in the network (or indirectly from national registries) to perform surveillance, or locally where participating centers in the network apply the algorithms for case-ascertainment themselves and share the outcome of surveillance with a coordinating center (for details, [22]). Local implementation allows for both semi and fully-AS, whereas centrally coordinated surveillance generally applies fully-AS.

For example, local implementation of semi-automated AS-HAI (surgical site infections) has been achieved in two surveillance networks, where an AS system for case ascertainment is implemented in multiple healthcare facilities, guided by coordinating centers to streamline efforts. Surveillance results are subsequently shared with the coordinating center [20, 26]. In two semi-automated AS-SARI examples, cases that result from queries on ICD-10 codes applied in healthcare facilities are subsequently shared with the coordinating center [27, 28].

Conversely, the Danish HAIBA system is an example of centrally implemented fully-automated AS-HAI, where nationally collected data on microbiology, admissions, procedures and antibiotic use are centrally applied to fully-AS [19]. Examples of a centrally implemented fully-automated AS-SARI are seen in multiple European countries, including Norway and Scotland, where ICD-10 codes and (sometimes anonymized) patient data are centrally collected from healthcare facilities or national registries [29, 30].

Experiences with local implementation of semi-AS have shown that large reductions in surveillance workload can be achieved without loss of sensitivity [31]. In addition, semi-automated AS-SARI systems allow for a more precise identification of SARI cases, as it can be enriched with information on symptoms, that is generally not available in EHRs in structured format. However, implementing AS has proven cumbersome [15, 18]. Local implementation of AS-HAI requires substantial knowledge of AS systems, increased IT capacity, and involvement of many different stakeholders and hence also demands advanced project management and maintenance (Brekelmans et al., manuscript submitted for publication published in ARIC: <https://doi.org/10.1186/s13756-024-01418-0>). For AS-SARI, an additional downside is that systems are not comprehensive but only cover the catchment population of selected healthcare facilities, which may not be representative on a national scale [32, 33] Another limitation of locally implemented

surveillance, from a central point of view, is the fact that external validation of locally implemented systems is limited, which brings uncertainty about the comparability across centers.

Although centrally implemented surveillance has advantages regarding the burden imposed on healthcare facilities, and consequently on sustainability and scalability, this strategy often leads to reduced timeliness, uncertainty about the catchment population (AS-SARI) and lack of detailed clinical information such as (severity of) symptoms and treatment. Diversity in documentation of routine care data and EHR systems hinder central implementation of fully AS and unstructured data are generally not applicable for reuse [30]. In addition, extensive linkage or sharing of data might be precluded by the General Data Protection Regulation (GDPR) in many countries [34], besides possible technical limitations. As a result, centrally implemented surveillance relies to a larger extent on diagnosis codes. Unfortunately, the wider use of ICD-10 codes in AS-SARI does not give the benefit of internationally comparable results, due to differences between countries in the selection of diagnosis codes and in case definitions [10]. Finally, the fully-automated and standardized AS occurring further away from the clinical environment brings its own challenges with respect to understanding and interpreting care processes and achieving useful feedback to clinicians. However, aligning surveillance methods with clinical practice is paramount to achieve optimal surveillance results and acceptance by clinicians [18, 21, 31].

In sum, both locally and centrally implemented AS are characterized by their own challenges, balancing the burden of local investment of resources and possibilities for variability against the more limited data availability in central implementation. The optimal solution would thus be decreased local burden and increased central and local oversight, while preserving privacy. As we will illustrate in the next section, the alternative approach of a federated infrastructure may be a solution to these needs.

A federated system as a potential alternative approach for implementation of automated surveillance

For AS, a federated system could be an alternative implementation strategy which is in between a centrally and locally implemented system. A federated system can be defined as an infrastructure where centrally developed analyses are applied locally (e.g. in healthcare facilities; Fig. 1). This approach thus facilitates the same analyses on all data, without the need for data sharing or central storage of source data. These analyses can be straightforward, such as a query for summary statistics (counts, proportions), but can also be more complex, like regression

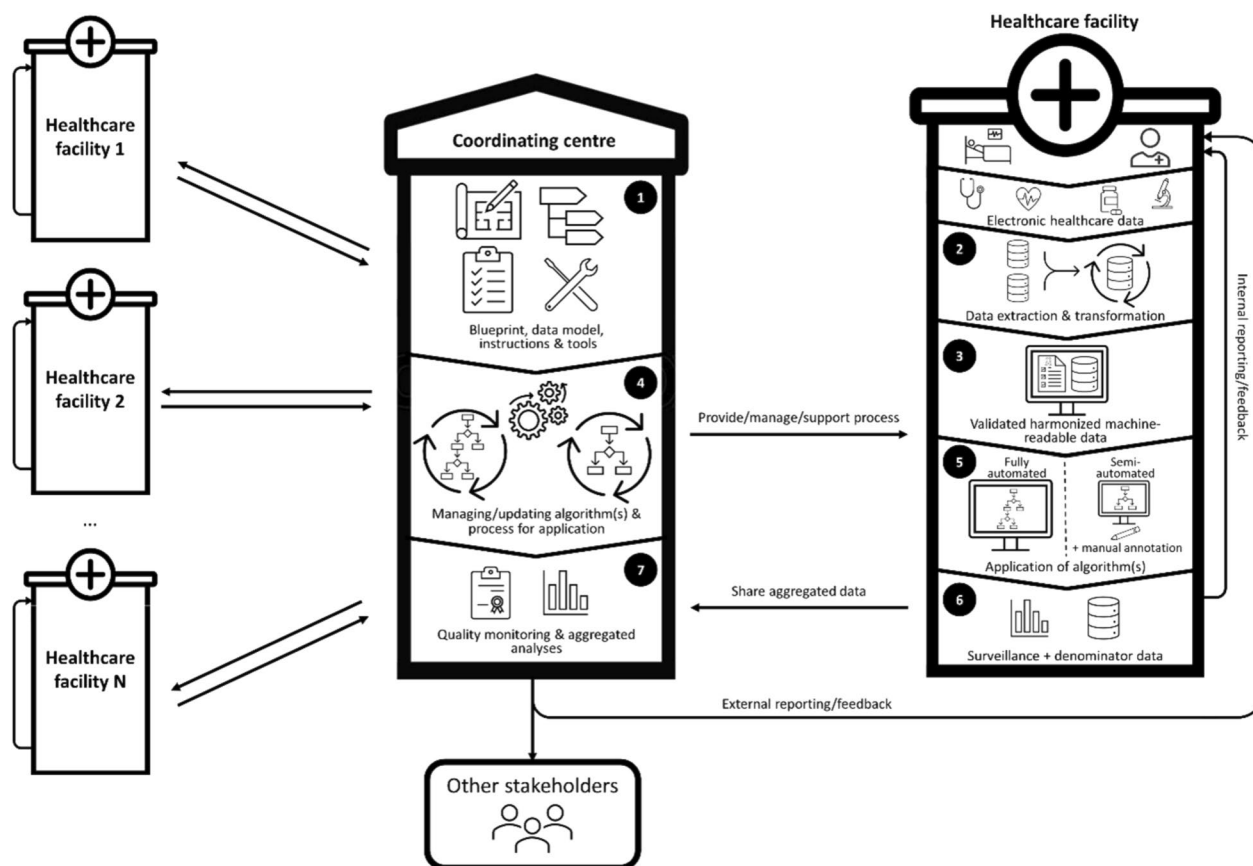


Fig. 1 Schematic representation of technical operationalization of federated AS. The coordinating centre designs, maintains and distributes all important aspects around the technical solution, algorithms and scripts for analyses, and data quality monitoring. Algorithms are locally applied and (aggregated) surveillance results are shared centrally. (Feedback) information can be shared with participating centres and other stakeholders in agreement with regulatory frameworks

or machine learning models. For AS, analyses can include algorithms for case identification. This is illustrated by an algorithm cloud service for HAI surveillance in the Patient Safety Surveillance Solutions System, which can be integrated in a federated system [35]. The exact architecture of a federated system will need to be tailored to the needs for surveillance and of stakeholder preferences. The details of the solution in practice will also depend on existing structures and processes in the environment where it is implemented.

Technical operationalization of federated AS

The coordinating centre develops a blueprint for design and maintenance of the technical solution (Fig. 1, point 1). The source data are locally extracted from EHR systems and subsequently transformed to harmonized machine readable (interoperable) data (explained in detail below (harmonization); Fig. 1, point 2) and validated (Fig. 1, point 3) When requested, centrally developed algorithms are run on these harmonized data at

different sites in the network (Fig. 1, point 4–5). Case-based or aggregated data can be shared and analysed on a regional, national, or international level, and results of these analyses can be fed back to the local sites and other stakeholders (Fig. 1, point 7). The coordinating centre provides management and support during the whole federated (semi-)AS process, including monitoring of (local) data quality and maintenance, distribution and validation of the algorithms. (Fig. 1, point 1,4,7) Governance aspects and prerequisites for federated AS will be discussed below.

Harmonization of data from electronic health records

Federated systems rely on harmonized digital routine care data from EHRs. Therefore, a clear understanding of key aspects of harmonization of data is important.

The digitalization of information in healthcare facilities has so far mainly targeted individual healthcare facility services rather than considering the overall process

within the organizations. As a result, information in healthcare facilities is often fragmented, collected in separate systems that do not interoperate (i.e. interact) and stored in different, in many cases proprietary, formats [36, 37]. Ideally, a common format is applied directly in primary EHR systems, but since such extensive standardization has not been reached yet information often needs data re-entry or transformation to be available across systems even within the same organization.

Data harmonization: semantical and syntactical interoperability

To facilitate the exchange of surveillance data within a surveillance network and enable the application of AS algorithms across different healthcare facilities, it is necessary that data across the different organizations share a common format and become interoperable. Interoperability has been defined as “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” [38].

We can identify four main levels of interoperability: technical, organizational, syntactic, and semantic. The first level refers to the capability of exchanging data using common communication protocols. Organizational interoperability alludes to the alignment of policies and goals across organizations. Syntactic and semantic interoperability focus on the format and on the terminology of the information to enable data harmonization and facilitate data exchange. These concepts are explained by practical examples.

In different healthcare facility systems, the same clinical concepts can be identified using different terms or different codes according to specific practice or conventions. For example, the microorganism “*Staphylococcus aureus*” could be identified with abbreviations such as “sau” “St. Au.” or other local codes across different systems. It is evident that this inhomogeneous terminology hinders automatic processing of data as each concept needs to be manually evaluated to understand its meaning and identify possible matches in the different systems.

To address this problem, international standard terminologies should be used; these provide unique codes that unequivocally identify clinical concepts at a global level. SNOMED CT [39], for example, is a comprehensive terminology for health-related terms that includes also international codes for the names of microorganisms, so that the code “3092008”, identifies unambiguously the *Staphylococcus aureus*.

Several widely used international terminologies with different focuses exist. Additional examples include the Logical Observation Identifiers Names and Codes (LOINC) [40] which allows to define laboratory procedures such as the detection of a microorganism

specifying different diagnostic methods; the Anatomical Therapeutic Chemical (ATC) [41] and the International Classification of Diseases (ICD) [42], endorsed by the World Health Organization, which provide standard codes for pharmaceutical products and diseases respectively. Within the project ORCHESTRA, standard terminologies were used to establish semantic interoperability across several COVID-19 studies in Europe [43] enabling federated analysis across the different data sources [44]. The Reconciliation of Cohort Data for Infectious Diseases (ReCoDID) Consortium harmonized and standardized clinical-epidemiological and laboratory data of nine arbovirus (arthropod-borne viruses) cohorts in Latin America, to reuse cohort data which individually may contain too small numbers to answer research questions [45]. Additionally, standards were also used to create common data elements for diagnostic tests in infectious disease studies [46].

Aside from ensuring that the semantics are standardized, the syntax also requires standardization to make data readable by machines and allow for data exchange. The international standard organization Health Level 7 (HL7) [47], has developed one of the most widely used messaging standards for the exchange of clinical data between systems in healthcare facilities, called HL7 V2. The innovative HL7 standard FHIR (Fast Healthcare Interoperability Resources [48]), represents an evolution from the previous HL7 standards, combining their features with the use of the latest web standards. FHIR enables efficient exchange of information and is complementary with the abovementioned standard terminologies which can be conveniently integrated into its structure. Figure 2 illustrates an example of the combined use of FHIR elements and standard terminologies to describe the detection of *Staphylococcus aureus* via microbial culture. The example is based on the data model (version 2024) for microbiology procedures described within the German Medical Informatics Initiative. In this context, a data set for the detection of a microorganism and for the definition of its characteristics was defined, describing the type of test (e.g. “culture”), whether there was a positive result (e.g. “detected”), and if so, the name of the organism (e.g. “*Staphylococcus aureus*”), based on international interoperability standards such as FHIR, SNOMED CT and LOINC [49].

Development of common data model for AS

To perform AS within a federated network of healthcare facilities, a common data model needs to be developed. To achieve a common data model for AS, first consensus is to be reached among the relevant scientific communities of medical and surveillance experts on data elements that are minimally required (i.e. minimal data set; MDS)

- **Observation.code:** 11475-1 Microorganism identified in Specimen by Culture (LOINC)
- **Observation.value:** 260373001 Detected (SNOMED CT)
- **Observation.component.value:** 3092008 *Staphylococcus aureus* (SNOMED CT)

Fig. 2 Example of three FHIR Observation elements, integrating terminology. The FHIR element Observation.code can be associated to a LOINC code to specify the type of observation. The FHIR element Observation.value is used to describe the result of the observation, in this case, using the SNOMED CT code for “detected”. The name of the detected micro-organism, “*Staphylococcus aureus*” in the example, can be defined in the FHIR element Observation.component.value specified by the corresponding SNOMED CT code

in the context of AS, for population selection, algorithm application for automated case detection, and case mix variables. Second, for this MDS, the semantic codes that identify the concepts need to be selected from the standard terminologies, accompanied with the specification of the syntax (e.g. in FHIR) in which the selected information should be provided. This second step requires a careful study of the relevant terms and international coding systems and presupposes a close collaboration between subject matter experts, information technology and healthcare standard experts.

This is not a one-off exercise, but maintenance is pivotal for high quality data. For example, new pathogens registered in laboratory information systems must be mapped to e.g. SNOMED CT codes or missing codes can be added [50]. It also requires coordination between healthcare facilities and the coordinating center: healthcare facilities need to actively inform the coordinating center when new elements are encountered. Conversely, the coordinating center needs to promptly inform healthcare facilities on version updates and data specifications.

We suggest that the common data model should be based on international interoperability standards to facilitate data compatibility across multiple applications at cross-institutional and also at international level. The use of standards reduces the need for data transformation and supports the process of making data Findable, Accessible, Interoperable, Reusable (FAIR) [51].

Prerequisites for implementation: local data transformation and quality assessment

Healthcare facilities participating in a network for federated AS thus need to extract the required data from EHRs, transform these data into the required harmonized and machine-readable formats (i.e. interoperable formats) and load them to establish the technical connection to make them accessible to be analyzed in a federated manner (Fig. 1, point 2). This process of transformation of data is known as Extract, Transform and Load (ETL) and requires IT experts from the healthcare facility to technically prepare the data to conform to the established common data model. These ETL processes

should be tailored to local information systems, be reproducible and performed repeatedly with different parameters, such as time periods. Depending on the surveillance target, this may require generation of real-time data, for example in detection of outbreaks or upsurges like in SARI surveillance. Establishing an efficient ETL process requires of course investment of resources, and it is therefore important to organize it in a structured way that can be re-utilized over time for different purposes. In Germany for example, the MII, has identified a core data set (MII CDS) of patient-related information that all university hospitals should be able to provide in a specific standard format for secondary use. As part of the initiative, hospitals have organized entities closely connected to the clinical information systems called “data integration centers” (DICs) where ETL processes make standardized MII CDS data available for many types of medical research or surveillance activities. DICs offer solutions for data exchange and analyses and are considered a crucial component of near future federative research projects, as well as fundamental for implementing automated HAI surveillance [52]. Data preparation is certainly not just a technical operation but requires also interpretation of the meaning of data. Source data from EHR systems have been primarily registered for healthcare provision, not for surveillance purposes, and have a meaning in a certain context. Local data transformation requires medical and domain experts with knowledge of the CDM and of local data. Because of differences in EHR and registration practices between healthcare facilities, mapping clinical information of the same concept to semantic codes is not a standard procedure. The level of detail of description of a disease, for example, may vary and consequently although the concept is the same, codes might be different [53]. For example, “Bacterial infectious disease (disorder)” and “Bacterial respiratory infection (disorder)” correspond to two different SNOMED codes as the latter concept includes information concerning the finding site for the disease; therefore, to maximize interoperability, general rules should be defined as to what level of information should be utilized. Likely, central support to local healthcare facilities by

providing knowledge and guidance in transformation of data and prepare for reuse, including supporting rules on the choice of terminology codes will be needed.

Validation is always vital for obtaining quality results and starts at the source of the data. Technical validation of data quality needs to be performed continuously within the ETL procedures and routine checks should assess whether the specifications of a CDM are met, if mandatory data are present and report outlier values. This is also not just a technical exertion, since medical and domain knowledge is required to interpret the data and make decisions on data selection, assessment of completeness and plausibility (e.g. expected gender, range of age) [54]. Therefore, close cooperation between the technicians who integrate the rules into the system and the domain experts who define the rules is necessary at central and local level.

In addition to data validation, it is important that algorithm results are validated locally. Even with data meeting specifications, good algorithm performance (e.g. sensitivity of case ascertainment) is not guaranteed, because of differences in clinical procedures. Since local algorithm validation is likely to be burdensome, it would be beneficial to investigate how sound validation can be performed without requiring a lot of resources and detailed knowledge on AS development, in line with the advantages that is strived for with federated AS.

Finally, validation also needs to be performed at a central level, both technical validation and content validation. For the latter, traceability and auditability of federated analyzed data is needed to ensure correct interpretation of the data [52, 55].

Prerequisites for implementation: ethical, legal and societal implications (ELSI) for application of federated automated surveillance

To better understand the potential ethical, legal and societal implications (the ELSI) of federated AS, an important question is how the GDPR [34] would apply to the processing of (personal) data through federated systems and if there are remaining concerns regarding data security, the privacy of persons, and other ELSI.

When focusing on implications for privacy and data protection, we assume that according to the federated systems as described above data are collected at the patient level. In this case, the same laws and regulations apply as in regular surveillance, including the GDPR. If shared data can be further aggregated, then data is considered non-personal data [56]. Still, there must be no available means to infer the identities of the underlying individuals from the non-personal data.

When considering compliance with the GDPR, the regulation delineates seven key principles [57]. The use of

federated systems clearly supports compliance with some of them, such as: purpose limitation (data are used for surveillance approved by provider), data minimization (only essential data are collected, preferably in aggregated or pseudorandomized form) and storage limitation (analyses on source data are performed locally). Four other principles are likely challenged, and need to be governed in the organization of a network for federated AS.

Firstly, concerning the principle of integrity & confidentiality, as mentioned, risks for exposure of personal data cannot be excluded, even in aggregated form. As such, additional data protection measures should be taken, such as through privacy enhancing techniques as described in the literature [58]. Secondly, as aforementioned, accuracy can be challenged as implementation of federated systems relies on substantial efforts from healthcare facilities and the coordinating centers to provide FAIR and quality data and to engage collaboratively in defining data standards. In addition, clear communications on specifications and validation procedures would align local processes to ensure comparability between healthcare facilities. Thirdly, accountability is complicated because responsibilities lie with different stakeholders (e.g. data quality from the provider side, versus data security from the user side) [59]. The surveillance process therefore necessitates a more collaborative character requiring a higher level of trust between all parties compared to traditional processing methods. Fourthly, lawfulness, fairness & transparency can be challenged by the “black-box” effect of algorithms and the federated system itself, which is less tangible and makes the exact data processing difficult to comprehend. In this regard, additional legal frameworks, such as the EU AI Act and the European Parliament’s resolution of 12 February 2020 on automated decision-making processes (2019/2915(RSP) [60, 61] become relevant addressing the risks to individuals from having automated tools used to inform decisions that might have an impact on them. In retrospective surveillance risks are considered low since results aims to inform policy decisions on a populational level and large and specified surveillance populations reduce the risk of sampling bias or unfair representation from populational groups. To conclude, stakeholder engagement and acceptance is paramount and could be achieved by involving them right from the design phase in development of federated AS, [62, 63] by creating clarity in expectations and optimizing usability, and by increasing reliance in the federated network by clear governance of responsibilities and accountability [64]. In addition, good communication is important to help address these challenges: not only between active participants of the federated AS systems to ensure better collaboration (e.g. communication on norms and standards to

enhance transparency and ensure data quality); but also to enhance the benefits of these systems at all levels [65].

Federated systems must comply with minimal transparency requirements that would allow users to make informed decisions. This underscores the need for regulatory frameworks for access, handling and publication of results, investing in digital literacy, as well as efforts from algorithm developers to make the source code transparent and understandable to healthcare facility professionals and the public, considering the final impact on quality of results and decision making [55, 63, 66]. Networks are to be established across all partners to build relationships, and feedback routes for defining protocols, shared aims and agendas, among other things [65]. It is important to reflect on how the relevant stakeholders, including civil society, would be properly informed on how federated analyses feed into decision-making. As an example, federated AS would allow national authorities, and individual hospitals to have a broader overview of the incidence of infections in their setting as compared to others together with communication on how these figures were arrived at. As such, it supports local decision making on preventive measures to improve such performance in the long term, without any impact on decisions about individual health care delivery. Supervision or independent oversight of AS by qualified professionals with legitimate public interests are critical, supported by risk management strategies and monitoring. Review structures should also have mechanisms where it is possible for stakeholders to question and remedy potential mistakes in decisions informed by automated processes.

As a final point, and from the perspective of a network on an international level, a challenge for complying with the GDPR, across all data analytics methods, is the difference between countries in interpretation and implementation of the legal framework, which makes it challenging to integrate datasets especially across different jurisdictions [67]. International networks of federated AS can help to address this issue by supporting the implementation of standards for regulatory compliance. Networks like VODAN-Africa included in their governance model, in addition to the well-known FAIR principles, the OLR (Ownership, Localisation and Regulatory Compliance) principles. The added value of this approach is that by adapting tools and practices to regulatory requirements across locations (usually based on the strictest ones) it creates a precedent enhancing not only legal certainty but also data interoperability, transparency, and creates opportunities for international collaborations [63, 68].

Pros and cons of federated systems for HAI and SARI surveillance

If we consider AS systems as the whole system of selection of surveillance population and source data, and identification of patients with SARI or HAI, and the aggregation of results federated AS appears to provide some major advantages over both locally and centrally implemented AS systems. Table 2 describes which needs for large-scale implementation of AS such a system would meet, but also highlights potential downsides and prerequisites for a sound methodology and sustainable implementation.

Federated AS combines some benefits of a centrally implemented system, such as standardization and alignment of an easily scalable centrally developed methodology, contributing to a reduction of the burden on individual healthcare facilities. At the same time, it offers some benefits of a locally implemented system, including more timely access to data and flexibility in algorithms and timing, satisfying different information needs and contributing to sustainability. Moreover, enabling semi-AS and more complex analyses methods would allow for a broader range of meaningful case definitions.

Federated AS systems may also prepare for application of technological advancements, such as 'federated learning'. This is a type of machine learning without central data collection. At every location, models are trained. Model metadata, rather than individual data points, are shared with the coordinating centre, for continued updating and refinement of shared global algorithms [55]. Although all analyses can be carried out with a central implementation strategy, it is often desirable to avoid sharing large amounts of data when possible, or in some cases sharing may even become unacceptable. In fact, the willingness to share data often is directly connected to the amount of (confidential) data being shared. Federated analyses could then enable more detailed analyses, that otherwise would have not been accepted in a non-federated setting. When we consider explanatory analyses for instance, we might develop a model based only on a limited number of covariates/features in a non-federated setting because of sharing concerns, whereas in a federated setting we might have a much larger amount of (sensitive) covariates/features to our disposal. An example of an analysis that could benefit from more detailed data is the identification of individual risk factors or risk factors at the level of the health care facility for the risk of HAI. Having more the data available then decreases the chances of overlooking any relevant risk factors. Next to more detailed analyses, a federated system might provide the opportunity to automate descriptive analyses, in which summary statistics are generated to provide an overview of the phenomena in regular time intervals

Table 2 Needs of large-scale automated healthcare facility surveillance and potential of federated systems

Current barriers precluding large-scale automated surveillance	Needs for large-scale automated infection surveillance in healthcare facilities	Potential benefits of federated systems for surveillance	Potential downsides of federated systems for surveillance	Prerequisites for federated systems for surveillance
Case-definitions are not always suitable for fully AS	Changes in definitions, centrally implemented semi-AS, or more complex case detection methods	Enables automated case detection in healthcare facilities by (complex) algorithms, supporting a broader range of definitions that may better align usefulness for local practitioners	Adaptability of algorithms to local situation more difficult to realize	Transparency and explainability of algorithms for case detection; sound methods for algorithm development and validation
Differences in registration procedures or routine care data limits comparability of case definitions across centers and over time	Uniformity when reusing EHR data (FAIR) ensures consistency of case-definition	Improved comparability of surveillance results due to harmonization of source data and improved insight in coding practices	Loss of information with data transformation; Knowledge and IT capacity for EHR data harmonization	Stakeholder agreement on a minimal data set, data model and terminology; Willingness and capacity for FAIR data curation;
Limited flexibility in information needs or granularity of existing national registries currently demands additional infrastructures or additional manual data collection (SARI)	Less resource intensive AS systems that are agile to meet current and possibly changing information needs	Sustainable systems that meet information needs for various purposes and topical issues by performance of various analyses on harmonized data sources at different time points, and different level of details	Methodology may not be accepted if data ownership and good governance are well arranged	Robust measures for quality control Regulatory framework for data access, handling and publication of results; Transparency of methods
Limited data linkage, due to limitations in interoperability or privacy restrictions	Authorized access to more detailed machine-readable data, allowing for less resource-intensive and more widely implemented surveillance systems	Potentially increased coverage rate of healthcare facilities (more complete information of catchment area); Privacy by design tailored to local situation; Methodology analyzing free text may be supported	Methodology may not be accepted unless validated and well understood; Lack of access to individual patient-level data on a central level; Data on rare events or combination of traits can make identifying individuals a risk, new information security issues or risk of hacking may not be excluded	Regulatory framework for data access, handling and publication of results; transparency in technical applications (freely accessible source code and pseudo code); Technical solution needs to fit the local needs and (regulatory) requirements; Thorough method validation
Insufficient capacity of IPC and IT in healthcare facilities for development, implementation, and maintenance of AS systems	Reduced workload for development, implementation, and maintenance of AS systems; Interoperable machine readable EHR data allowing exchange of algorithms and scripts	Surveillance systems programmed and maintained at one location; Easily scalable solution with interoperable, machine-readable data	Software ETL processes necessary every healthcare facility; accountability more complicated; IT capacity at the expense of IP capacity if the importance and shift in IP tasks is not recognized	Technical applications need to be transparent (open source); Governance framework addressing accountability, enhanced collaboration
Insufficiently broad and thorough knowledge on AS in all participating healthcare facilities hinders development or procurement of methodologically sound and sustainable AS systems	Central support in the development of methodologically sound and sustainable AS systems; easier applicable from AS by a large multidisciplinary team within the healthcare facilities	Programme of requirements of a sustainable AS system can be determined collectively within federated community; healthcare facilities are not dependent on individual agreements with vendors	Knowledge in all participating healthcare facilities is required to understand, accept and validate outcomes of federated AS. Potential loss of view on local events without manual assessment	Digital literacy and minimal training in data intelligence for understanding (maintenance of) algorithms and interpretation of outcomes, for decision making; supervision and oversight of AS methodology by qualified professionals

Table 2 (continued)

Current barriers precluding large-scale automated surveillance	Needs for large-scale automated infection surveillance in healthcare facilities	Potential benefits of federated systems for surveillance	Potential downsides of federated systems for surveillance	Prerequisites for federated systems for surveillance
External validation of locally implemented AS systems difficult (HA)	Source codes and data handling procedures transparent and easy to review	Surveillance algorithms programmed, validated and maintained at one location	Continuous quality monitoring, maintenance and version control to be performed by coordinating center	Framework for validation; transparency in data processing through whole pipeline from curation to publication of results for quality assurance, and avoidance of gaming in surveillance

(e.g. description of upsurges of influenza-positives before hospital admission, differences in epidemic severity between seasons or appearance of new pathogens in certain regions). Although Federated Learning methodology seems promising, few studies explored this new analysis method on EHR data [69, 70]. Challenges of federated learning related to the choice of analyses methods and models, and, by definition, their ability to provide reliable answers despite not having the metadata of models from all healthcare facilities available at the same time, have been identified in specific settings [71, 72]. And, despite the elegance of federated analyses, new security issues cannot be excluded [73]. The feasibility, reliability, and factors impacting the outcomes of federated learning need to be explored in the context of analyses relevant to HAI and SARI surveillance to be accepted and considered trustworthy by stakeholders [74].

Cross-institutional agreements on harmonization of source data would bring great advantages to healthcare facilities as they would be empowered to re-use the data for multiple applications and could easily join existing networks on a case-by-case basis. Harmonization of source data based on interoperability standards enhances interoperability and FAIRness, improves the comparability of surveillance results over time between healthcare facilities, and if data models are shared on an international national level, between surveillance networks. For example, in AS-SARI surveillance, the sensitivity of (proxy) case definitions depends on the specification of the source data [10]. Ideally, CDMs based on interoperability standards should be supported by international organizations such as European Centre for Disease Prevention and Control or other relevant international surveillance-related institutes to facilitate same source data specifications across various surveillance networks, allowing for comparability of surveillance results between networks and re-use of knowledge. This perfectly aligns with the WHO recommendations for resilient surveillance for respiratory viruses, indicating that data management systems and standards should be adopted for simultaneous and coordinated interpretation of data from multiple sources and countries are pivotal for respiratory illness surveillance to quickly triangulate information. This would allow for cross-country comparisons, potentially revealing essential information for decision making that would not have become visible based on surveillance in single countries [75].

Although with federated AS there is no need to develop, implement and maintain an AS system in each healthcare facility, data transformation and maintenance of IT connections still places a significant burden on local experts and also resources and a shift in knowledge necessary for meaningful interpretation should be considered. From a

study in German healthcare facilities, it was learned that data availability and accessibility from the EHR systems for reuse for surveillance purposes was variable with room for improvement, and adherence to interoperability standards would be strategic [76]. A federated system can only be implemented if the resources are planned carefully and sustainably, and are recommended not to be a single-purpose investment but have the flexibility to be also supportive for surveillance of other diseases and research. Implementing standards such as SNOMED CT or LOINC can be a challenging process that is generally not prioritized in the routine of laboratories and healthcare facility. However, the developments on harmonization and full exploitation of the potential of health data by making data FAIR is not isolated. Implementation is likely to be facilitated by (inter)national developments such as the European Health Data Space (EHDS) regulation [77, 78], aiming to solve the problem of segmentation of healthcare information by creating a framework for health data exchange and integration in Europe by providing rules and governance. These developments can be important for adoption of federated systems for surveillance.

While data may be harmonized, it should not be overlooked that the meaning or interpretation of data may differ. Differences between healthcare facilities may be the result of differences in coding practices or clinical policies, e.g. regarding diagnostic testing or medication prescriptions. The quality of data and validity of standardized algorithms has yet to prove itself, and opportunities for local adaptation may need to be explored. Moreover, data integrity and data quality become more difficult to assess and manage centrally, because cleaning and harmonization occur locally. No thorough knowledge on AS systems is needed in every healthcare facility, but a clear understanding of federated AS and (complex) algorithms is still required to interpret and accept the surveillance results. This underlines the importance of digital and health literacy to make informed decisions, but may also warrant good governance and organization, to facilitate healthcare facilities. Transparency and clear understanding of data are key when developing federated AS, to create valuable and actionable data supporting infection control measures.

The barriers and needs and potential benefits in this perspective stem mainly from experiences with large-scale implementation of automated surveillance. Reported barriers to implementing digital health solutions for public health surveillance and sharing data for analyses in lower resourced settings include obstacles to digitisation and manual efforts to collect, structure and submit data, resulting in loss of (integrity of) data and divergence on (quality) standards and legal frameworks

and regulations, and uncertainties or lack of ownership [63, 79, 80]. Well-designed federated automated surveillance, where aforementioned governance aspects, FAIR standards and architecture are included directly from the design phase, could address many of the abovementioned barriers. It thus may be beneficial to support less resourced settings to set up automated surveillance, on the premise that there is minimal digitalization of information and resources are sustained. In addition, through this approach, international and cross-sectorial collaborations could be enhanced supporting global public health and pandemic preparedness [12, 65, 75].

Conclusions, future perspectives

In conclusion, AS in federated systems can benefit from central development of systems for fully or semi-AS for large-scale implementation of AS-HAI and AS-SARI, without the need to share large amounts of source data, as a solution for current challenges and preparing for future developments. In addition, using harmonized source data could allow for increased comparability of results. Still, the quality of data and AS results must be proven. Additionally, further research is required to assess the impact of potential disadvantages of federated systems, including quantifying the required resources for data harmonization and validation and defining the governance structures and regulatory frameworks that should be in place.

We have now evaluated the potential of a federated approach in the context of specific surveillance aims. AS-HAI and AS-SARI share commonalities, such as the importance of uniformity of definitions and case detection methods and reliability of applied methods over time, but have also different requirements such as the scale of the population of interest and importance of timeliness. A considerable advantage of federated surveillance may be that the same infrastructure can be leveraged to scale up to other surveillance targets, such as surveillance of antimicrobial resistance, and in a wider global context to optimize care and mitigate public health threats (80). Ultimately, the method should prove itself through good quality surveillance data that are useful and actionable for local decision making to reach the goal, being prevention of infections.

Glossary

Fully-automated surveillance: Automated surveillance where all steps of surveillance - from data collection to the determination of the infection status are performed without any human intervention or interpretation [22].

Semi-automated surveillance: Automated surveillance where the determination of an infection status is a combination of automation, chart review and manual confirmation for a selection of

records [22].

Case-definition: An objective description of established criteria to identify the disease under surveillance

Source data: Raw data elements from routine care data used by algorithms to detect (possible) HAI, calculate the denominator or risk factors (e.g. microbiology results, admission and discharge dates, central line days, procedure codes) [22].

Algorithms for case identification: An automated detection method to classify cases with (a probability of) an infection according to the case definition

Centrally implemented automated surveillance: Automated surveillance designed, coordinated, and implemented the coordinating center (central case ascertainment) [22]

Locally implemented automated surveillance: Automated surveillance designed and coordinated by a coordinating center, but implemented locally (local case ascertainment), under the responsibility of the participating healthcare facility [22]

Aggregated data/results: Data reported on a group level as opposed to on an individual level

Federated system: An infrastructure that can be used for collaborative analyses across data from various local facilities, without sharing or central storage of these data

Federated analyses: Collaborative analyses across data from different institutes (e.g. healthcare facilities), without sharing source data

Federated learning: Type of machine learning where models are trained on local datasets that are not collected centrally,

Interoperability: The ability of two or more systems or components to exchange information and to use the information that has been exchanged

Common data model: Specifications on syntax and semantics for a specified data set with the purpose to facilitate interoperability and data exchange among separate applications and data sources

Interoperability standards: Specifications provided by Standard Development Organizations of the syntax and semantics of (health) data

Minimal data set: Data elements required for a specific surveillance purpose

ETL; Extract transform load: The process required to combine data from different sources into a specified format

Acknowledgements

We would like to thank Per Englund for his valuable contribution during discussions.

Author contributions

All authors (S.R., S.W., M.M., F.L., K.L.M., S.V., C.S.R., A.W., S.H., M.B., E.R.) contributed to drafting the manuscript and performing a critical review.

Funding

No funding was received for this perspective.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Competing interests

Suzanne Ruhe van der Werff is a shareholder of the company P3S (Patient Safety Surveillance Solutions).

Author details

¹Department of Epidemiology and Surveillance, Centre for Infectious Disease Epidemiology and Surveillance, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. ²Division of Infectious Diseases, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. ³Department of Infectious Diseases, Karolinska University Healthcare Facility, Stockholm, Sweden. ⁴Department of Medical Microbiology and Infection Control, University Medical Centre Utrecht, Utrecht, The Netherlands. ⁵Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Copenhagen, Denmark. ⁶Center for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. ⁷Department of Statistics Data Science en Modelling, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. ⁸Central Diagnostic Laboratory, University Medical Centre Utrecht, Utrecht, The Netherlands. ⁹Institute of Hygiene and Environmental Medicine, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin and, Berlin Institute of Health, Berlin, Germany. ¹⁰National Reference Center for the Surveillance of Nosocomial Infections, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health, Berlin, Germany. ¹¹Core Unit Digital Medicine and Interoperability, Berlin, Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany.

Received: 29 May 2024 Accepted: 8 September 2024

Published online: 27 September 2024

References

- Centers for Disease Control and Prevention (CDC). Introduction to Public Health Surveillance [Available from: <https://www.cdc.gov/training/publichealth101/surveillance.html>] Accessed April 26, 2024.
- World Health Organization. Surveillance [Available from: <https://www.who.int/emergencies/surveillance>. Accessed 07 Apr 2024.
- 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. <https://doi.org/10.1186/s13756-016-0149-9>.
- Price L, MacDonald J, Melone L, Howe T, Flowers P, Currie K, et al. Effectiveness of national and subnational infection prevention and control interventions in high-income and upper-middle-income countries: a systematic review. *Lancet Infect Dis*. 2018;18(5):e159–71. [https://doi.org/10.1016/S1473-3099\(17\)30479-6](https://doi.org/10.1016/S1473-3099(17)30479-6).
- Suetens C, Latour K, Karki 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. <https://doi.org/10.2807/1560-7917.ES.2018.23.46.1800516>.
- World Health Organization. Global report on infection prevention and control. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO. World Health Organization; 2022.
- Tomczyk S, Twyman A, de Kraker MEA, Coutinho Rehse AP, Tartari E, Toledo JP, et al. The first WHO global survey on infection prevention and control in health-care facilities. *Lancet Infect Dis*. 2022;22(6):845–56. [https://doi.org/10.1016/S1473-3099\(21\)00809-4](https://doi.org/10.1016/S1473-3099(21)00809-4).
- Rosenthal VD, Yin R, Nercelles P, Rivera-Molina SE, Jyoti S, Dongol R, et al. International nosocomial infection control consortium (INICC) report of health care associated infections, data summary of 45 countries for 2015 to 2020, adult and pediatric units, device-associated module. *Am J Infect Control*. 2024;52(9):1002–11. <https://doi.org/10.1016/j.ajic.2023.12.019>.
- Fakih MG, Bufalino A, Sturm L, Huang RH, Ottenbacher A, Saake K, et al. Coronavirus disease 2019 (COVID-19) pandemic, central-line-associated bloodstream infection (CLABSI), and catheter-associated urinary tract infection (CAUTI): The urgent need to refocus on hardwiring prevention efforts. *Infect Control Hosp Epidemiol*. 2022;43(1):26–31. <https://doi.org/10.1017/ice.2021.70>.
- Pelat C, Bonmarin I, Ruello M, Fouillet A, Caserio-Schonemann C, Levy-Bruhl D, et al. Improving regional influenza surveillance through a combination of automated outbreak detection methods: the 2015/16 season in France. *Euro Surveill*. 2017. <https://doi.org/10.2807/1560-7917.ES.2017.22.32.30593>.
- van't Klooster TM, Wielders CC, Donker T, Isken L, Meijer A, van den Wijngaard CC, et al. Surveillance of hospitalisations for 2009 pandemic influenza A(H1N1) in the Netherlands 5 June - 31 December 2009. *Euro Surveill*. 2010. <https://doi.org/10.2807/ese.15.02.19461-en>.
- Gupta S, Gupta T, Gupta N. Global respiratory virus surveillance: strengths, gaps, and way forward. *Int J Infect Dis*. 2022;121:184–9. <https://doi.org/10.1016/j.ijid.2022.05.032>.
- Birgand G, Lepelletier D, Baron G, Barrett S, Breier AC, Buke C, et al. Agreement among healthcare professionals in ten European countries in diagnosing case-vignettes of surgical-site infections. *PLoS ONE*. 2013;8(7):e68618. <https://doi.org/10.1371/journal.pone.0068618>.
- Mitchell BGH, Halton K, MacBeth D, Gardner A. Time spent by infection control professionals undertaking healthcare associated infection surveillance: A multi-centred cross sectional study. *Infect. Dis Health*. 2016;21(1):36–40. <https://doi.org/10.1016/j.idh.2016.03.003>.
- Shenoy ES, Branch-Elliman W. Automating surveillance for healthcare-associated infections: rationale and current realities (Part I/III). *Antimicrob Steward Healthc Epidemiol*. 2023;3(1):e25. <https://doi.org/10.1017/ash.2022.312>.
- Sips ME, Bonten MJM, van Mourik MSM. Automated surveillance of healthcare-associated infections: state of the art. *Curr Opin Infect Dis*. 2017;30(4):425–31. <https://doi.org/10.1097/QCO.0000000000000376>.
- Lin MY, Woeltje KF, Khan YM, Hota B, Doherty JA, Borlowsky TB, et al. Multicenter evaluation of computer automated versus traditional surveillance of hospital-acquired bloodstream infections. *Infect Control Hosp Epidemiol*. 2014;35(12):1483–90. <https://doi.org/10.1086/678602>.
- Verberk JDM, van der Kooij TIL, Hetem DJ, Oostdam N, Noordergraaf M, de Greeff SC, et al. Semiautomated surveillance of deep surgical site infections after colorectal surgeries: a multicenter external validation of two surveillance algorithms. *Infect Control Hosp Epidemiol*. 2023;44(4):616–23. <https://doi.org/10.1017/ice.2022.147>.
- Statens Serum Institut. Healthcare-Associated Infections Database (HAIBA) [Available from: <https://miba.ssi.dk/overvaagningssystemer/haiba>. Accessed April 26, 2024.
- Picard J, Nkoumazok B, Arnaud I, Verjat-Trannoy D, Astagneau P. Comorbidities directly extracted from the hospital database for adjusting SSI risk in the new national semiautomated surveillance system in France: The SPICMI network. *Infect Control Hosp Epidemiol*. 2023. <https://doi.org/10.1017/ice.2023.123>.
- 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(6):970–6. <https://doi.org/10.1093/cid/cix835>.
- 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(Suppl 1):S3–19. <https://doi.org/10.1016/j.cmi.2021.02.028>.
- National Healthcare Safety Network (NHSN) CDC. Ventilator-associated Events (VAE) [Available from: <https://www.cdc.gov/nhsn/psc/vae/index.html>. Accessed 26 Apr 2024.
- World Health Organization. Surveillance case definitions for ILI and SARI [Available from: <https://www.who.int/teams/global-influenza-programme/surveillance-and-monitoring/case-definitions-for-ili-and-sari#:~:text=SARI%20case%20definition,within%20the%20last%2010%20days>. Accessed 26 Apr 2024
- Cauchi JP, Borg ML, Dziugyte A, Attard J, Melillo T, Zahra G, et al. Digitalizing and upgrading severe acute respiratory infections surveillance in Malta: system development. *JMIR Public Health Surveill*. 2022;8(12):e37669. <https://doi.org/10.2196/37669>.
- Dutch National Institute for Public Health and the Environment (RIVM). PREZIES Automatisering Surveillance: POWI ORTHOpédie (PAS ORTHO) [Available from: <https://www.rivm.nl/prezies/pas-ortho>. Accessed 27 Nov 2023
- Buda S, Tolksdorf K, Schuler E, Kuhlen R, Haas W. Establishing an ICD-10 code based SARI-surveillance in Germany—description of the system and first results from five recent influenza seasons. *BMC Public Health*. 2017;17(1):612. <https://doi.org/10.1186/s12889-017-4515-1>.
- Torres AR, Gomez V, Kislaya I, Rodrigues AP, Fernandes Tavares M, Pereira AC, et al. Monitoring COVID-19 and influenza: the added value of a severe

- acute respiratory infection surveillance system in Portugal. *Can J Infect Dis Med Microbiol.* 2023;2023:6590011. <https://doi.org/10.1155/2023/6590011>.
29. Wells J, Young JJ, Harvey C, Mutch H, McPhail D, Young N, et al. Real-time surveillance of severe acute respiratory infections in Scottish hospitals: an electronic register-based approach, 2017–2022. *Public Health.* 2022;213:5–11. <https://doi.org/10.1016/j.puhe.2022.09.003>.
 30. Whittaker R, Toikkanen S, Dean K, Lyngstad TM, Buanes EA, Klovstad H, et al. A comparison of two registry-based systems for the surveillance of persons hospitalised with COVID-19 in Norway, February 2020 to May 2022. *Euro Surveill.* 2023. <https://doi.org/10.2807/1560-7917.ES.2023.28.33.2200888>.
 31. van Rooden SM, Tacconelli E, Pujol M, Gomila A, Kluytmans J, Romme J, et al. A framework to develop semiautomated surveillance of surgical site infections: an international multicenter study. *Infect Control Hosp Epidemiol.* 2020;41(2):194–201. <https://doi.org/10.1017/ice.2019.321>.
 32. Brady M, Duffy R, Domegan L, Salmon A, Maharjan B, O'Broin C, et al. Establishing severe acute respiratory infection (SARI) surveillance in a sentinel hospital, Ireland, 2021 to 2022. *Euro Surveill.* 2023. <https://doi.org/10.2807/1560-7917.ES.2023.28.23.2200740>.
 33. Fischer N, Dauby N, Bossuyt N, Reynders M, Gerard M, Lacor P, et al. Monitoring of human coronaviruses in Belgian primary care and hospitals, 2015–20: a surveillance study. *Lancet Microbe.* 2021;2(3):e105–14. [https://doi.org/10.1016/S2666-5247\(20\)30221-4](https://doi.org/10.1016/S2666-5247(20)30221-4).
 34. General Data Protection Regulation. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>.
 35. Patient Safety Surveillance Solutions (P3S) [Available from: <https://www.p3s.se/problem-solution>. Accessed 26 Aug 2024.
 36. Grüttner P. Opening the door for digital transformation in hospitals: IT expert's point of view. In: Glauner P, Plugmann P, Lerzynski G, editors. *Digitalization in healthcare: implementing innovation and artificial intelligence*. Cham: Springer International Publishing; 2021. p. 29–42.
 37. Lehne M, Sass J, Essenwanger A, Scheepers J, Thun S. Why digital medicine depends on interoperability. *NPJ Digit Med.* 2019;2:79. <https://doi.org/10.1038/s41746-019-0158-1>.
 38. Gansel X, Mary M, van Belkum A. Semantic data interoperability, digital medicine, and e-health in infectious disease management: a review. *Eur J Clin Microbiol Infect Dis.* 2019;38(6):1023–34. <https://doi.org/10.1007/s10096-019-03501-6>.
 39. SNOMED International. Available from: <https://www.snomed.org/>. Accessed 26 Apr 2024
 40. LOINC. Available from: <https://loinc.org/>. Accessed 26 Apr 2024
 41. Norwegian Institute for Public Health, WHO Collaborating Centre for Drug Statistics Methodology. International language for drug utilization research [Available from: <https://atcddd.fhi.no/>. Accessed 26 Apr 2024
 42. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision. Available from: <https://icd.who.int/browse/10/2019/en>. Accessed 26 Apr 2024
 43. Rinaldi E, Stellmach C, Rajkumar NMR, Caroccia N, Dellacasa C, Giannella M, et al. Harmonization and standardization of data for a pan-European cohort on SARS-CoV-2 pandemic. *NPJ Digit Med.* 2022;5(1):75. <https://doi.org/10.1038/s41746-022-00620-x>.
 44. Dellacasa C, Ortali M, Rossi E, Abu Attieh H, Osmo T, Puskaric M, et al. An innovative technological infrastructure for managing SARS-CoV-2 data across different cohorts in compliance with general data protection regulation. *Digit Health.* 2024;10:20552076241248920. <https://doi.org/10.1177/20552076241248920>.
 45. Gomez G, Hufstedler H, Montenegro Morales C, Roell Y, Lozano-Parra A, Tami A, et al. Pooled cohort profile: ReCoDID Consortium's harmonized acute febrile illness arbovirus meta-cohort. *JMIR Public Health Surveill.* 2024;10: e54281. <https://doi.org/10.2196/54281>.
 46. Stellmach C, Hopff SM, Jaenisch T, Nunes de Miranda SM, Rinaldi E, Napkon LO, ReCo DIDWG. Creation of standardized common data elements for diagnostic tests in infectious disease studies: semantic and syntactic mapping. *J Med Internet Res.* 2024;26: e50049. <https://doi.org/10.2196/50049>.
 47. Health Level Seven International (HL7). [Available from: <https://www.hl7.org/>. Accessed 26 Apr 2024.
 48. Health Level Seven International (HL7). HL7 FHIR Specification [Available from: <https://hl7.org/fhir/directory.html>. Accessed 26 Apr 2024.
 49. Rinaldi E, Drenkhahn C, Gebel B, Saleh K, Tonnie H, von Loewenich FD, et al. Towards interoperability in infection control: a standard data model for microbiology. *Sci Data.* 2023;10(1):654. <https://doi.org/10.1038/s41597-023-02560-x>.
 50. Hagel S, Gantner J, Spreckelsen C, Fischer C, Ammon D, Saleh K, et al. Hospital-wide electronic medical record evaluated computerised decision support system to improve outcomes of Patients with staphylococcal bloodstream infection (HELP): study protocol for a multicentre stepped-wedge cluster randomised trial. *BMJ Open.* 2020;10(2): e033391. <https://doi.org/10.1136/bmjopen-2019-033391>.
 51. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016;3: 160018. <https://doi.org/10.1038/sdata.2016.18>.
 52. Prokosch HU, Gebhardt M, Gruendner J, Kleinert P, Buckow K, Rosenau L, Semler SC. Towards a national portal for medical research data (FDPG): vision, status, and lessons learned. *Stud Health Technol Inform.* 2023;302:307–11. <https://doi.org/10.3233/SHTI230124>.
 53. Martinez-Costa C, Abad-Navarro F. Towards a semantic data harmonization federated infrastructure. *Stud Health Technol Inform.* 2021;281:38–42. <https://doi.org/10.3233/SHTI210116>.
 54. Behnke M, Valik JK, Gubbels S, Teixeira D, Kristensen B, Abbas M, et al. Information technology aspects of large-scale implementation of automated surveillance of healthcare-associated infections. *Clin Microbiol Infect.* 2021;27(Suppl 1):S29–39. <https://doi.org/10.1016/j.cmi.2021.02.027>.
 55. Raab R, Kuderle A, Zakreuskaya A, Stern AD, Klucken J, Kaissis G, et al. Federated electronic health records for the European Health Data Space. *Lancet Digit Health.* 2023;5(11):e840–7. [https://doi.org/10.1016/S2589-7500\(23\)00156-5](https://doi.org/10.1016/S2589-7500(23)00156-5).
 56. Podda E. Shedding light on the legal approach to aggregate data under the GDPR & the FFDR. CONFERENCE OF EUROPEAN STATISTIANS; Expert Meeting on Statistical Data Confidentiality; 2021–12–1. Poland2021. https://unece.org/sites/default/files/2021-12/SDC2021_Day1_Podda_AD.pdf.
 57. European Commission. Principles of the GDPR [Available from: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr_en. Accessed 26 Apr 2024.
 58. Maurya JP, S. Privacy Preservation in Federated Learning: its Attacks and Defenses. 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN); Salem India. 2023. <https://doi.org/10.1109/ICPCSN58827.2023.00177>
 59. Lo SK, Liu Y, Qinghua L, Wang C, Xiwei X, Paik H-Y, Zhu L. Toward trustworthy AI: blockchain-based architecture design for accountability and fairness of federated learning systems. *IEEE Internet Things J.* 2023;10(4):3276–84. <https://doi.org/10.1109/JIOT.2022.3144450>.
 60. European Artificial Intelligence Act, P9_TA(2024)0138 (2024). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html.
 61. European Union Automated decision-making processes: Ensuring consumer protection, and free movement of goods and services; European Parliament resolution of 12 February 2020 on automated decision-making processes: ensuring consumer protection and free movement of goods and services (2019/2915(RSP)) 2021 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020IP0032>.
 62. Bezemer T, de Groot MC, Blasse E, Ten Berg MJ, Kappen TH, Bredenoord AL, et al. A human(e) factor in clinical decision support systems. *J Med Internet Res.* 2019;21(3): e11732. <https://doi.org/10.2196/11732>.
 63. van Reisen M, Oladipo F, Stokmans M, Mpezamihgo M, Folorunso S, Schultes E, et al. Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Adv Genet (Hoboken).* 2021;2(2): e10050. <https://doi.org/10.1002/ggn2.10050>.
 64. Kerasidou CX, Kerasidou A, Buscher M, Wilkinson S. Before and beyond trust: reliance in medical AI. *J Med Ethics.* 2022;48(11):852–6. <https://doi.org/10.1136/medethics-2020-107095>.
 65. World Health Organization. Defining collaborative surveillance: a core concept for strengthening the global architecture for health emergency preparedness, response, and resilience (HEPR). Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO.: World Health Organization; 2023.
 66. Corbucci L, Guidotti R, Monreale A. Explaining black-boxes in federated learning. In: Longo L, editor. *Explainable artificial intelligence: first world*

- conference, xAI 2023, Lisbon, Portugal, July 26–28, 2023, Proceedings, Part II. Cham: Springer Nature Switzerland; 2023. p. 151–63. https://doi.org/10.1007/978-3-031-44067-0_8.
67. European Parliamentary Research Service (STOA) SFU How the General Data Protection Regulation changes the rules for scientific research 2019 [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU\(2019\)634447\(ANN1\)_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634447/EPRS_STU(2019)634447(ANN1)_EN.pdf).
 68. van Reisen M, Amare SY, Nalugala R, Taye GT, Gebreselassie TG, Medhanyie AA, Schultes E, Mpezamihigo M. Federated FAIR principles: Ownership, localisation and regulatory compliance (OLR). *FAIR Connect*. 2023;1(1):63–9. <https://doi.org/10.3233/FC-230506>.
 69. Pan W, Xu Z, Rajendran S, Wang F. An adaptive federated learning framework for clinical risk prediction with electronic health records from multiple hospitals. *Patterns (N Y)*. 2024;5(1): 100898. <https://doi.org/10.1016/j.patter.2023.100898>.
 70. Rajendran S, Xu Z, Pan W, Ghosh A, Wang F. Data heterogeneity in federated learning with Electronic Health Records: Case studies of risk prediction for acute kidney injury and sepsis diseases in critical care. *PLOS Digit Health*. 2023;2(3): e0000117. <https://doi.org/10.1371/journal.pdig.0000117>.
 71. Li W, Tong J, Anjum MM, Mohammed N, Chen Y, Jiang X. Federated learning algorithms for generalized mixed-effects model (GLMM) on horizontally partitioned data from distributed sources. *BMC Med Inform Decis Mak*. 2022;22(1):269. <https://doi.org/10.1186/s12911-022-02014-1>.
 72. Zhu H, Zhang H, Jin Y. From federated learning to federated neural architecture search: a survey. *Complex Intell Syst*. 2021;7:639–57. <https://doi.org/10.1007/s40747-020-00247-z>.
 73. Shokri R, Stronati M, Song C, Shmatikov V. Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy; 2017. pp. 3–18. <https://doi.org/10.1109/SP.2017.41>.
 74. Hand D. Trustworthiness of Statistical Inference. *J R Stat Soc A Stat Soc*. 2022;185(1):329–47. <https://doi.org/10.1111/rssa.12752>.
 75. World Health Organization. “Crafting the mosaic” a framework for resilient surveillance for respiratory viruses of epidemic and pandemic potential. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. World Health Organization; 2023.
 76. Aghdassi SJS, Goodarzi H, Gropmann A, Clausmeyer J, Geffers C, Piening B, et al. Surgical site infection surveillance in German hospitals: a national survey to determine the status quo of digitalization. *Antimicrob Resist Infect Control*. 2023;12(1):49. <https://doi.org/10.1186/s13756-023-01253-9>.
 77. European Commission Recommendation on a European Electronic Health Record exchange format 2019 <https://digital-strategy.ec.europa.eu/en/library/recommendation-european-electronic-health-record-exchange-format>.
 78. European Parliament Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space 2022 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197>.
 79. Karamagi HC, Muneene D, Droti B, Jepchumba V, Okeibunor JC, Nabyonga J, et al. eHealth or e-Chaos: the use of digital health interventions for health systems strengthening in Sub-Saharan Africa over the last 10 years: a scoping review. *J Glob Health*. 2022;12:04090. <https://doi.org/10.7189/jogh.12.04090>.
 80. Kozlakidis Z, Abduljawad J, Al Khathaami AM, Schaper L, Stelling J. Global health and data-driven policies for emergency responses to infectious disease outbreaks. *Lancet Glob Health*. 2020;8(11):e1361–3. [https://doi.org/10.1016/S2214-109X\(20\)30361-2](https://doi.org/10.1016/S2214-109X(20)30361-2).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.