Data-Driven Surveillance: Effective Collection, Integration, and Interpretation of Data to Support Decision Making

Fernanda C. Dórea[1]* and Crawford W. Revie[2]

[1] Department of Disease Control and Epidemiology, National Veterinary Institute, Uppsala, Sweden, [2] Computer and Information Sciences, University of Strathclyde, Glasgow, United Kingdom

The biggest change brought about by the "era of big data" to health in general, and epidemiology in particular, relates arguably not to the volume of data encountered, but to its variety. An increasing number of new data sources, including many not originally collected for health purposes, are now being used for epidemiological inference and contextualization. Combining evidence from multiple data sources presents significant challenges, but discussions around this subject often confuse issues of data access and privacy, with the actual technical challenges of data integration and interoperability. We review some of the opportunities for connecting data, generating information, and supporting decision-making across the increasingly complex "variety" dimension of data in population health, to enable data-driven surveillance to go beyond simple signal detection and support an expanded set of surveillance goals.

Keywords: epidemiology, machine learning, big data, data analyses, linked data

INTRODUCTION

Increases in data volume, diversity and speed have affected all aspects of human life. As we advance into the 21st century, Simonsen et al. (1) highlight two main streams that are pushing health surveillance into the "Big Data Era": the advancements in laboratorial detection tools which traditional surveillance rely on, and a dramatic increase in the number of health and non-health related data streams that can be exploited for surveillance. However, as Leyens et al. (2) point out, "the simple fact that there is more data is not useful to public health unless we are able to turn it into 'actionable data' for improved health outcomes and more effective and efficient health systems."

While health surveillance systems continue to adapt, improving traditional components [e.g., (2–4)] and adding others based on the exploitation of novel data streams [e.g., (5–8)], their progress fades in comparison to that seen in other sectors (1), from business and marketing to the more related area of diagnostic services within human health. While data scientists seem to agree that a significant big data trend in 2017 was an end to talk about it as if it were a novelty (9), in health surveillance "big data" remains a buzz word. A number of publications have discussed the challenges and potential benefits of incorporating big data into surveillance, but a framework for the operationalization of data-driven surveillance has seldom been discussed. Moreover, discussions around the exploitation of novel data streams has been focused almost exclusively on emergence prediction and early disease detection, in detriment of other surveillance goals, such as situational awareness for non-communicable and endemic diseases, and disease freedom demonstration.

Based on the results of a workshop carried out in late 2017, and supported by a scoping review, we discuss the challenges and opportunities for implementing data-driven surveillance frameworks as a 3-step process: data integration; data processing to generate information; and making outputs from data analyses accessible and usable by decision-makers.

METHODS

On October 10th and 11th, 2017, the Uppsala Heath Summit gathered around 200 delegates from different sectors, and from around the world, to discuss priorities for preventing, detecting and responding to infectious disease threats using a One Health approach (10). A dedicated 3 h workshop was conducted by the authors to explore the theme of innovation and big data in health surveillance. The 63 workshop participants brainstormed to identify and prioritize opportunities to achieve data-driven decision-making in population health, within the One Health context. Participants came from a range of sectors: 16 were from universities, 11 from the private sector, 22 from governmental agencies and one from a global health organization. This was a multi-disciplinary group, from the fields of public health (11), animal health (12), pharmacovigilance (13), health and medicine (3), data science (4), climate (1), and geography (1). Most participants worked in European countries, with three participants from Africa, two from North America and one from South America. Informed by a literature search targeting articles in the health surveillance domain which used the term "big data," workshop discussions were organized into four main groups of "big data analytics" (BDA) challenges: technical, operational, normative (cultural and ethical challenges), and funding. A summary of the workshop discussions, within the four main challenge themes, is already available in the post-conference report (10). Following the workshop, we have organized the discussion according to actual implementation steps, laying out a "data to actionable information" continuum, and enriched it with bibliography relevant for each section.

We have also updated and reviewed the literature search specifically targeting BDA. We searched Scopus for papers published up to December 2020 in the general area of health surveillance which contained the term "big data" [TITLE-ABS-KEY ("big data" AND surveillance AND (health OR disease OR syndromic))]. This search returned 492 papers. After reviewing title and abstract, and reading selected papers for which full-text was available in English, we selected a total of 47 papers which specifically discuss data science and data innovation challenges and opportunities in any area of health surveillance.

We have not cited all papers here due to space limitations, but the full list of 44 selected papers is available in the **Supplementary Material**, and also at (http://datadrivensurveillance.org/dds_ICAHS2020).

RESULTS AND DISCUSSION
Step 1—Connecting Data
The most significant changes in the area of health data in general, and epidemiology in particular, arguably relate not to the volume of data, but to their variety. An increasing number of innovative data sources, including many not collected specifically for health purposes, can now be used for epidemiological inference and contextualization (14). The challenges of data integration have been discussed by many researchers (2, 12, 13, 15, 16). Often, however, the discussion confuses issues around data access and privacy, with the actual technical challenges of data integration and interoperability. The latter issues are central to contemporary surveillance, which increasingly relies on combining evidence from multiple data sources.
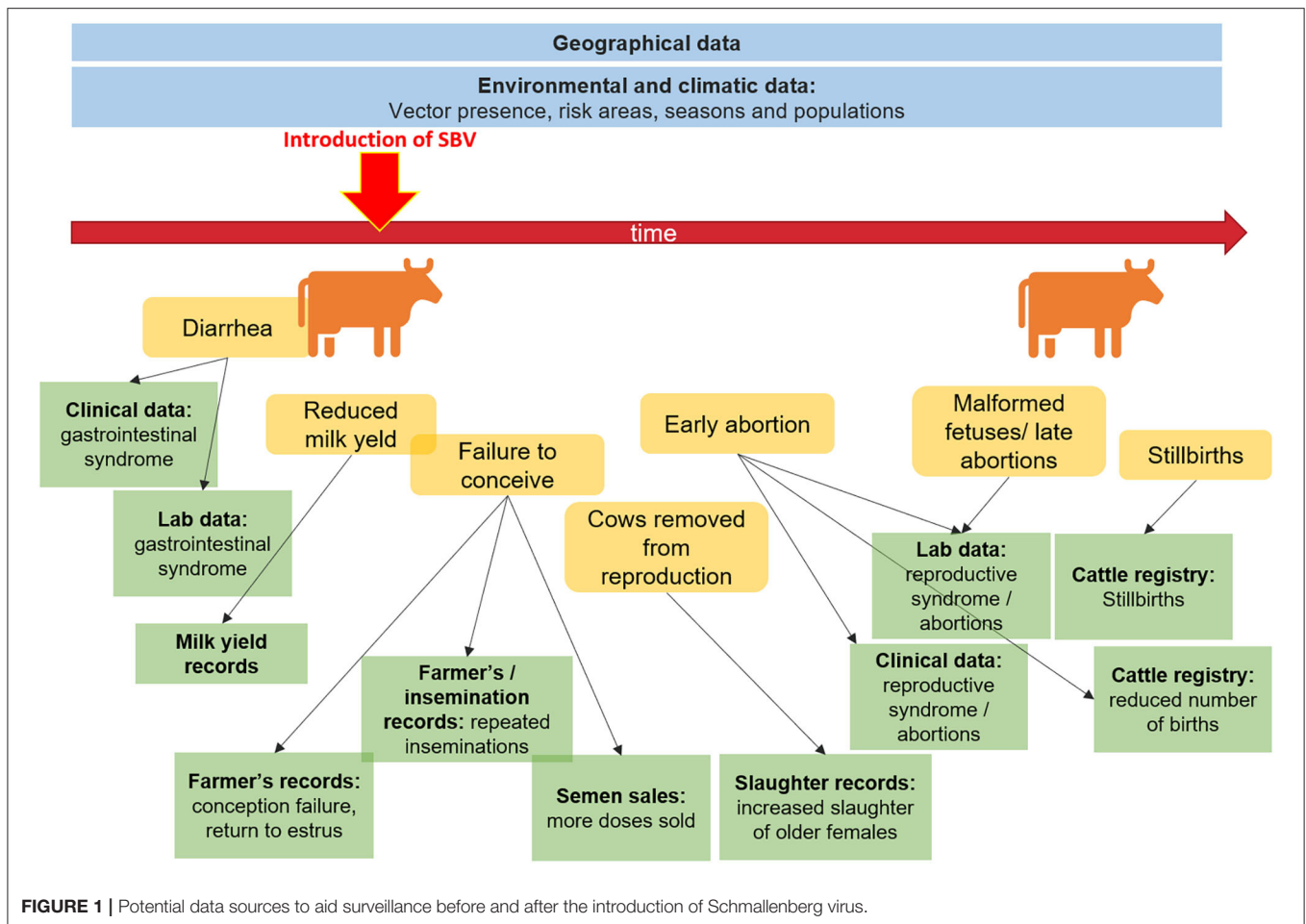
Surveillance data have traditionally been classified by mode of acquisition: active or passive. With the advent of "big data," the concept of data acquisition becomes less central—we move from intentionally producing surveillance data, to taking advantage of ubiquitous data sources generated as a part of many processes, health related, or not (11). The technical challenge is no longer validating a dataset in which each observation was intentionally recorded, but rather mining data streams for valid evidence to support decision making (11).

Figure 1 illustrates the potential data streams from which signals of a health hazard occurrence might originate for the case of Schmallenberg virus being introduced into a dairy herd. This figure represents a limited snapshot of the health continuum of interest for animal health. We can imagine the increased complexity involved if we were to consider a zoonotic pathogen, and had to factor in exposure to humans through the food production cycle, or environmental exposure. The variety of novel data streams that can support surveillance has been reviewed in detail for animal health (17, 18), drug safety and health care (2), food safety (19), and one medicine (20). Opportunities associated specifically with spatial data sources (21) and search query data (22) have also been reviewed.

In addition to the access and interpretation of a greater number of opportunistic data sources, there are also increased opportunities to redesign the purposeful collection of surveillance data in the digital era. Salathé (14) discussed applications to drug safety monitoring, while a broader review of crowd-sourcing, citizen sensing and sensor web technologies for health is given by Kamel Boulos et al. (23). Workshop participants highlighted, in particular, the use of apps for patient reporting or self- diagnosis, which can have value along the entire surveillance continuum: from prevention, to communication with the public during response.

The sources of data we have access to determine the types of evidence we can extract, and the timeliness of such extraction. As Han and Drake (24) note, our ability to move toward predictive capacity is limited not by technology, but by access to appropriate data. To achieve a paradigm shift in disease control, moving from disease response to disease intelligence, a resilient health system must be underpinned by environmental, geographic, and population data (2, 24).

During the workshop, the group concluded that the single biggest barrier to gaining insights from data, particularly in real-time settings, was data integration. The need to "break the barriers of siloed data" was often mentioned as a priority. Timely access to integrated data was considered the main challenge

FIGURE 1 | Potential data sources to aid surveillance before and after the introduction of Schmallenberg virus.

to using data-driven evidence in emergencies, such as during outbreak response.

The issue of data integration and interoperability (25) is particularly important when targeting long chains involving multiple actors, such as in food safety surveillance (19). The lack of standardized data was repeatedly mentioned as a barrier for data processing and interpretation. However, as the discussion around this issue matured, most participants agreed that it was unrealistic to expect data standardization, as in fact many standards already exist for health data, but are not used. Most importantly, many existing standards contribute only to achieving structural (syntactic) interoperability.

As the secondary use of data sources (re-use) increases, and models demand integration of data from multiple disciplines, we will increasingly require semantic interoperability. Semantic interoperability is concerned with ensuring that the integrity and meaning of the data is preserved throughout the integration process (26). This is achieved by storing data in machine interoperable formats making use of knowledge models that explicitly document, for humans and for machines, the domain knowledge and assumptions under which data were collected and are stored (27). Ontologies allow domain

experts to create knowledge models that can be interpreted both by humans and machines (28). Using such models, computers can reason with data without relying on the use of specific codification. For an example in animal health, see Dórea et al. (29).
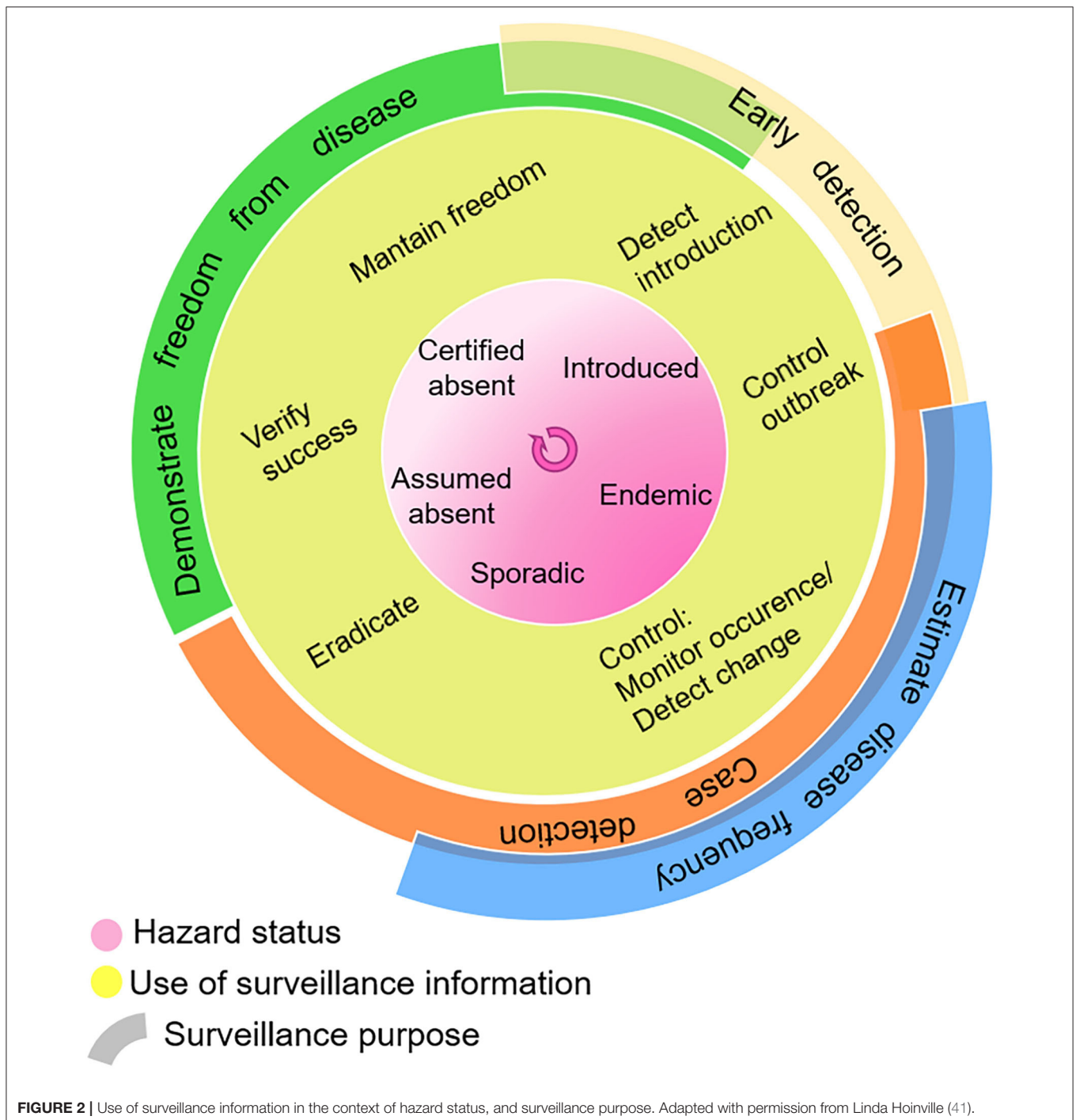
Step 2—Generating Information

A common skepticism related to big data comes from authors who highlight its potential to become a "hypothesis generating machine," capable of detecting correlation, but not causation (12, 30). The question should perhaps not be whether big data are useful, but what they are useful for. In surveillance, associations may be an important source of information for decision on interventions that aim at risk mitigation or case finding, even in the absence of any proven causal association. Iwashyna and Liu (11) point out that the questions which big data cannot answer are similar to those that are also a challenge in most observational studies, such as prescriptive questions. The authors suggest three main types of questions that can be addressed with big data: prognostic questions (what is going to happen), which "require temporally stable associations, not underlying causal models"; predictive questions (what will likely

happen if something different is done); and patterning questions (describing population patterns).

Automated access to continuous streams of data has allowed monitoring of population patterns—and early detection of unexpected changes—at earlier and earlier steps on the disease continuum. From direct monitoring of early registers (e.g., veterinarian calls or visits to the emergency room), to even less specific, but earlier signs of health change, such as over-the-counter drug sales. This component has been coined "syndromic surveillance" due to the initial focus on the monitoring of unspecific clinical symptoms in public health (31). The methodology has been applied in animal health to a number of data sources that are not necessarily "syndromic" (32, 33), and its utility is being increasingly explored for situational awareness

FIGURE 2 | Use of surveillance information in the context of hazard status, and surveillance purpose. Adapted with permission from Linda Hoinville (41).

rather than simply for early disease detection. To that end, Smith et al. (34) reported the need to focus on system sustainability and usefulness as one of the main lessons learned from two decades experience with syndromic surveillance in the UK. They argued that systems should be designed with a focus on the uses, not the data sources, and should aim to serve multiple public health objectives.
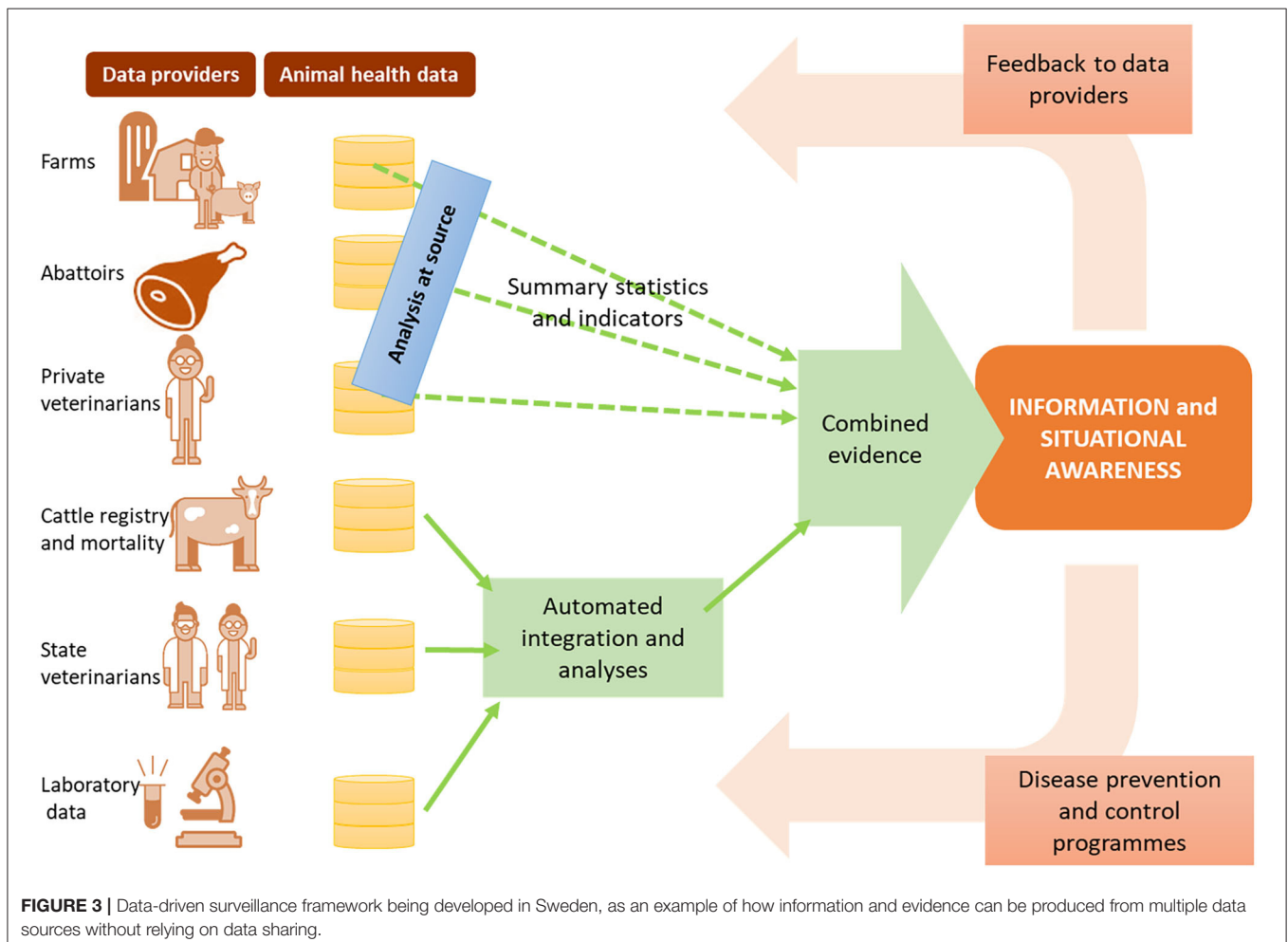
For a more complete review of the architectures and specific methods for big data analysis in health surveillance, we refer readers to (19, 25, 35, 36). For a review of the use of terms "big data," "informatics" and "bioinformatics" in the animal health and veterinary medical literature, we refer to (37).

During the workshop, the discussion focused not on what analysis tools to use, but on how to incorporate available methods within routine surveillance. The gap between technological and methodological innovation, as well as implementation in field settings are also discussed in (38). An important message related to the fact that surveillance officials should not only have access to the right tools, but should also be capable of using them effectively. "Efficient people and technology," as one captured note summarized this point. The need for more training was repeatedly listed, in addition to the importance of making tools that are more accessible to domain experts; that is, user-friendly and available in local languages. Chiolero and Buckeridge (39) called these the "knowledge brokers" needed to "bridge data science, health monitoring and public health." Reference was also made to the training needs discussion presented in Brownson et al. (40).

Step 3—Supporting Decision-Making

Surveillance activities are designed according to the desired use of surveillance information, as summarized in **Figure 2**. This in turn depends on the hazard occurrence in the target population or geographical area. As can also be seen in this figure, the boundaries are not always clear, and purposes can overlap. This highlights an overall workshop conclusion that the separation of surveillance goals may be artificial, and that a data-driven decision support system should be designed to strengthen all stages of disease control. Chiolero and Buckeridge (39) emphasize the role of decision-makers in identifying surveillance needs, setting priorities, and evaluating the effect of interventions. They added to their "glossary of public health surveillance in the era of big data" the idea of a continuum from data, to information, to *evidence* (which "emerges from

FIGURE 3 | Data-driven surveillance framework being developed in Sweden, as an example of how information and evidence can be produced from multiple data sources without relying on data sharing.

the comparison of information"), then "used to build actionable knowledge" (DIEK pyramid) (42).

Increases in data variety and velocity have opened up new surveillance opportunities, most notably in relation to disease prevention and early detection. The ability to train statistical algorithms on a large quantity and variety of data to identify relationships and monitor interactions allows us to monitor risks in space and time [creating a "riskscape" (24)], and respond to these risks, rather than to occurrence. It creates the opportunity to improve timeliness and population coverage, and increase resolution (spatial and temporal) (25), leading to infectious disease intelligence—knowing what, when, why, and how to respond (24). In public health, the use of new data and technologies to assess population health with increased accuracy and granularity at temporal and geographical levels, delivering programs tailored to specific populations, has been coined "precision public health" (39, 43).

While the advent of "big data analysis" has been extensively discussed for disease prediction and early response, its support to other surveillance goals has often been overlooked. Access to digitalized and novel data streams can increase the timeliness of surveillance information, but can also "improve temporal or spatial resolution of surveillance, add surveillance to places with no existing systems,... measure aspects of a transmission/disease process not captured by traditional surveillance, and increase the population size under surveillance" (44). Antoine-Moussiaux et al. (45) argue that a focus on detection of disease signals may miss the true value of surveillance, which lies in its continuity. They propose that health surveillance should be viewed as an information system, which continuously provides feedback to inform the prioritization of actions.

This assumes we have addressed the two previous steps, and as such have access not simply to "big data," but to FAIR data—findable, accessible, interoperable and reusable (46). In a scenario of semantically interoperable data we can more readily employ machines to reason over complex knowledge, and support surveillance decision-making holistically. Data variety and even issues of data accessibility are resolved, rather than being barriers. In an ongoing project in Sweden, for example (**Figure 3**), we are researching methods to combine evidence from analysis, rather than combining data directly. Data are analyzed at source, with signals being compiled centrally.

Moreover, a data-driven surveillance framework assumes that decision-makers have access to the outputs of big data analysis with the same level of "FAIRness" —this requires the availability of decision supporting dashboards that allow end users to query through the data sources in consumable formats, and navigate through the outputs of analysis in transparent ways. Most importantly, it requires that the value extracted from the data is returned to all relevant stakeholders (**Figure 3**), creating a positive cycle of encouragement not only for data accessibility, but also for data quality.

CONCLUSION

Solving the technological barriers to extracting information from big data is only the first step toward a framework for evidence-based decision making. Data-driven support to surveillance in practice will depend on having access to the right data, employing the right methods, and making the outputs accessible and understandable to the right stakeholders. Participants in the workshop, as well as several papers reviewed (1, 14, 19, 47), highlighted that data-driven components could support traditional surveillance, but that the surveillance systems of the future will be a hybrid of traditional and data-driven methods. System design should focus on health surveillance goals and utility to the decision-makers. Information generation is data-driven, but system design should not be. Using novel data sources to complement those used traditionally will merge the best of both worlds—though gains in timeliness and predictive power will come at the cost of dealing with all of the complexity in these novel data sources (1).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fvets. 2021.633977/full#supplementary-material

REFERENCES

1. Simonsen L, Gog JR, Olson D, Viboud C. Infectious disease surveillance in the big data era: toward faster and locally relevant systems. *J Infect Dis.* (2016) 214 (Suppl. 4):S380–5. doi: 10.1093/infdis/jiw376

2. Leyens L, Reumann M, Malats N, Brand A. Use of big data for drug development and for public and personal health and care. *Genet Epidemiol.* (2017) 41:51–60. doi: 10.1002/gepi.22012

3. Pfeiffer DU, Stevens KB. Spatial and temporal epidemiological analysis in the Big Data era. *Prev Vet Med.* (2015) 122:213–20. doi: 10.1016/j.prevetmed.2015.05.012

4. Lee ECCEC, Asher JMMJM, Goldlust S, Kraemer JDJDD, Lawson ABABB, Bansal S. Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. *J Infect Dis.* (2016) 214 (Suppl. 4):S409–13. doi: 10.1093/infdis/jiw344

5. Gogoshin G, Boerwinkle E, Rodin AS. New algorithm and software (BNOmics) for inferring and visualizing bayesian networks from heterogeneous big biological and genetic data. *J Comput Biol.* (2017) 24:340–56. doi: 10.1089/cmb.2016.0100

6. Kumar ATK, Asamoah D, Sharda R. *Can Social Media Support Public Health? Demonstrating disease surveillance using big data analytics.* Puerto Rico: Americas Conference on Information Systems, AMCIS (2015).

7. O'Shea J, O'Shea J, O'Shea J. Digital disease detection: a systematic review of event-based internet biosurveillance systems. *Int J Med Inform.* (2017) 101:15–22. doi: 10.1016/j.ijmedinf.2017.01.019

8. Dórea FCFC, Sanchez J, Revie CWCW. Veterinary syndromic surveillance: current initiatives and potential for development. *Prev Vet Med.* (2011) 101:1–17. doi: 10.1016/j.prevetmed.2011.05.004

9. Mayo M. *Big Data: Main Developments in 2016 and Key Trends.* (2017). Available online at: http://www.kdnuggets.com/2016/12/big-data-main-developments-2016-key-trends-2017.html (accessed November 11, 2019).

10. "Uppsala Health Summit." *Tackling Infectious Disease Threats: Prevent, Detect, Respond with a One Health Approach.* In: Conclusions from Uppsala Health Summit 10–11 October (2017). Uppsala (2017). Available online at: https://www.uppsalahealthsummit.se/digitalAssets/547/c_547164-l_1-k_uppsala-health-summit-post-conference-report-2017.pdf

11. Iwashyna TJ, Liu V. What's so different about big data? A primer for clinicians trained to think epidemiologically. *Ann Am Thorac Soc.* (2014) 11:1130–5. doi: 10.1513/AnnalsATS.201405-185AS

12. Hoffman S, Podgurski A. Big bad data: law, public health, and biomedical databases. *J Law, Med Ethics.* (2013) 41(Suppl. 1):56–60. doi: 10.1111/jlme.12040

13. Asokan GV, Asokan V. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. *J Epidemiol Glob Health.* (2015) 5:311–4. doi: 10.1016/j.jegh.2015.02.001

14. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. Bourne PE, editor. *PLoS Comput Biol.* (2012) 8:1–5. doi: 10.1371/journal.pcbi.1002616

15. Vayena E, Salathé M, Madoff LC, Brownstein JS. Ethical challenges of big data in public health. *PLoS Comput Biol.* (2015) 11:e1003904. doi: 10.1371/journal.pcbi.1003904

16. Toh S, Platt R. Is size the next big thing in epidemiology? *Epidemiology.* (2013) 24:349–51. doi: 10.1097/EDE.0b013e31828ac65e

17. Gates MC, Holmstrom LK, Biggers KE, Beckham TR. Integrating novel data streams to support biosurveillance in commercial livestock production systems in developed countries: challenges and opportunities. *Front Public Heal.* (2015) 3:74. doi: 10.3389/fpubh.2015.00074

18. VanderWaal K, Morrison RB, Neuhauser C, Vilalta C, Perez AM. Translating big data into smart data for veterinary epidemiology. *Front Vet Sci.* (2016) 4:110. doi: 10.3389/fvets.2017.00110

19. Marvin HJP, Janssen EM, Bouzembrak Y, Hendriksen PJM, Staats M. Big data in food safety: an overview. *Crit Rev Food Sci Nutr.* (2017) 57:2286–95. doi: 10.1080/10408398.2016.1257481

20. McCue ME, McCoy AM. The scope of big data in one medicine: unprecedented opportunities and challenges. *Front Vet Sci.* (2017) 4:1–23. doi: 10.3389/fvets.2017.00194

21. Stevens KB, Pfeiffer DU. Sources of spatial animal and human health data: casting the net wide to deal more effectively with increasingly complex disease problems. *Spat Spatiotemporal Epidemiol.* (2015) 13:15–29. doi: 10.1016/j.sste.2015.04.003

22. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. *PLoS ONE.* (2014) 9:e109583. doi: 10.1371/journal.pone.0109583

23. Kamel Boulos MN, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, et al. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *Int J Health Geogr.* (2011) 10:67. doi: 10.1186/1476-072X-10-67

24. Han BA, Drake JM. Future directions in analytics for infectious disease intelligence: toward an integrated warning system for emerging pathogens. *EMBO Rep.* (2016) 17:e201642534. doi: 10.15252/embr.201642534

25. Bansal S, Chowell G, Simonsen L, Vespignani A, Viboud C. Big data for infectious disease surveillance and modeling. *J Infect Dis.* (2016) 214:S375–9. doi: 10.1093/infdis/jiw400

26. Definition of Interoperability. *HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations.* 2nd ed. Chicago, IL: HIMSS (2010). p. 190.

27. Noy N. Semantic integration: a survey of ontology-based approaches. *SIGMOD Rec.* (2004) 33:65–70. doi: 10.1145/1041410.1041421

28. Noy NF, McGuinness DL. *Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory.* (2001) Available online at: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

29. Dórea FC, Vial F, Hammar K, Lindberg A, Lambrix P, Blomqvist E, et al. Drivers for the development of an Animal Health Surveillance Ontology (AHSO). *Prev Vet Med.* (2019) 166:39–48. doi: 10.1016/j.prevetmed.2019.03.002

30. Khoury MJMJMJ, Ioannidis JPAJPA. Medicine. Big data meets public health. *Science.* (2014) 346:1054–5. doi: 10.1126/science.aaa2709

31. Centers for Disease Control and Prevention (CDC). Syndromic surveillance. Reports from a national conference. *Morb Mortal Wkly Rep.* (2004) 53 (Suppl.):1–264.

32. Dórea FC, Vial F. Animal health syndromic surveillance: a systematic literature review of the progress in the last 5 years (2011–2016). *Vet Med Rep.* (2016) 7:157–69. doi: 10.2147/VMRR.S90182

33. Dupuy C, Bronner A, Watson E, Wuyckhuise-Sjouke L, Reist M, Fouillet A, et al. Inventory of veterinary syndromic surveillance initiatives in Europe (Triple-S project): current situation and perspectives. *Prev Vet Med.* (2013) 111:220–9. doi: 10.1016/j.prevetmed.2013.06.005

34. Smith GE, Elliot AJ, Lake I, Edeghere O, Morbey R, Catchpole M, et al. Syndromic surveillance: Two decades experience of sustainable systems—Its people not just data! Epidemiol Infect. (2019) 147:e101. doi: 10.1017/S0950268819000074

35. Huang T, Lan L, Fang X, An P, Min J, Wang F. Promises and challenges of big data computing in health sciences. *Big Data Res.* (2015) 2:2–11. doi: 10.1016/j.bdr.2015.02.002

36. Wang W, Krishnan E. Big data and clinicians: a review on the state of the science. *JMIR Med Inform.* (2014) 2:e1. doi: 10.2196/medinform.2913

37. Ouyang Z, Sargeant J, Thomas A, Wycherley K, Ma R, Esmaeilbeigi R, et al. A scoping review of "big data", "informatics", and "bioinformatics" in the animal health and veterinary medical literature. *Anim Heal Res Rev.* (2019) 1–18. doi: 10.1017/S1466252319000136

38. Buckee C. Improving epidemic surveillance and response: big data is dead, long live big data. *Lancet Digit Heal.* (2020) 2:e218–20. doi: 10.1016/S2589-7500(20)30059-5

39. Chiolero A, Chiolero A, Chiolero A, Chiolero A, Buckeridge D. Glossary for public health surveillance in the age of data science. *J Epidemiol Community Health.* (2020) 74:612–6. doi: 10.1136/jech-2018-211654

40. Brownson RC, Samet JM, Bensyl DM. Applied epidemiology and public health: are we training the future generations appropriately? *Ann Epidemiol.* (2017) 27. doi: 10.1016/j.annepidem.2016.12.002

41. RISKSUR surveillance design framework. Available online at: https://survtools.org/wiki/surveillance-design-framework/ (accessed November 11, 2020).

42. Dammann O. Data, information, evidence, and knowledge: a proposal for health informatics and data science. *Online J Public Health Inform.* (2019) 10:e224. doi: 10.5210/ojphi.v10i3.9631

43. Dolley S. Big data's role in precision public health. *Front Public Heal.* (2018) 6:681. doi: 10.3389/fpubh.2018.00068

44. Althouse BM, Scarpino S V., Meyers LA, Ayers JW, Bargsten M, Baumbach J, et al. Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Sci.* (2015) 4:1–8. doi: 10.1140/epjds/s13688-015-0054-0

45. Antoine-Moussiaux N, Vandenberg O, Kozlakidis Z, Aenishaenslin C, Peyre M, Roche M, et al. Valuing health surveillance as an information system: interdisciplinary insights. *Front Public Heal.* (2019) 7:138. doi: 10.3389/fpubh.2019.00138

46. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18

47. Kostkova P. *A Roadmap to Integrated Digital Public Health Surveillance: The Vision and the Challenges.* In: *Companion - Proceedings of the 22nd International Conference on World Wide Web.* Rio de Janeiro (2013). doi: 10.1145/2487788.2488024