Biosurveillance enterprise for operational awareness, a genomic-based approach for tracking pathogen virulence

Willy A Valdivia-Granda

Orion Integrated Biosciences Inc.; New Rochelle, NY USA

To protect our civilians and warfighters against both known and unknown pathogens, biodefense stakeholders must be able to foresee possible technological trends that could affect their threat risk assessment. However, significant flaws in how we prioritize our countermeasure-needs continue to limit their development. As recombinant biotechnology becomes increasingly simplified and inexpensive, small groups, and even individuals, can now achieve the design, synthesis, and production of pathogenic organisms for offensive purposes. Under these daunting circumstances, a reliable biosurveillance approach that supports a diversity of users could better provide early warnings about the emergence of new pathogens (both natural and manmade), reverse engineer pathogens carrying traits to avoid available countermeasures, and suggest the most appropriate detection, prophylactic, and therapeutic solutions. While impressive in data mining capabilities, real-time content analysis of social media data misses much of the complexity in the factual reality. Quality issues within free-form user-provided hashtags and biased referencing can significantly undermine our confidence in the information obtained to make critical decisions about the natural vs. intentional emergence of a pathogen. At the same time, errors in pathogen genomic records, the narrow scope of most databases, and the lack of standards and interoperability across different detection and diagnostic devices, continue to restrict the multidimensional biothreat assessment. The fragmentation of our biosurveillance efforts into different approaches has stultified attempts to implement any new foundational enterprise that is more reliable, more realistic and that avoids the scenario of the warning that comes too late. This discussion focus on the development of genomic-based decentralized medical intelligence and laboratory system to track emerging and novel microbial health threats in both military and civilian settings and the use of virulence factors for risk assessment. Examples of the use of motif fingerprints for pathogen discrimination are provided.

Correspondence to: Willy A Valdivia-Granda;
Email: willy.valdivia@orionbio.com

Introduction

Transcontinental migratory dynamics and economic exchange of commodities have resulted in an increasing exposure of humans to new infectious diseases and the circulation into urban areas of zoonotic pathogens previously found only in tropical, remote or unpopulated locations. Due to military conflict and humanitarian relief efforts, our civilian and military forces are deployed into areas characterized by uncertainty and complexity, where encounters with endemic pathogens can affect their operational readiness. Furthermore, the unique characteristics of many bacteria, virus, and toxins, coupled with progress in genetic engineering and synthetic biology techniques, have opened new dimensions in regards to the potential development of bioweapons with enhanced infectivity, virulence, vaccine avoidance, and antimicrobial resistance.[1-3] This situation is compounded by the scale and speed of technological developments with no historic precedent that continues to profoundly affect life-sciences technologies of dual-use. Unlike nuclear weapons, which are both difficult and expensive to build, deadly pathogens are quickly becoming inexpensive to modify, design, develop, produce, and use for biowarfare and bioterrorism.[4,5] The challenges of this situation will represent difficult strategic and tactical issues for both senior civilian leadership and military commanders.

In the past 12 years, the US government has established a framework to regulate the possession, transfer, and reporting of pathogens that can disrupt global health. Many of these initiatives emerged after the American anthrax attacks and are grounded in the resulting 2004 Homeland Security Presidential Directives 10 (HSPD-10), which established four biodefense pillars: (1) threat awareness, (2) prevention and protection, (3) surveillance and detection, and (4) response and recovery. These directives instructed the identification of vulnerabilities and recommended the creation of a national bio-awareness system. The 2007 HSPD-21 directive named biosurveillance as a critical priority for improving public health and instructed all appropriate government branches to propose strategies for tracking and reporting pathogens. The 2010 National Strategy for Countering Biothreats (NSCB) and the 2012 National Strategy for Biosurveillance (NSB) directives consolidated the need for obtaining timely and accurate insights on current and emerging pathogens and presented the research community with yet another policy framework for pathogen detection,

characterization, and reporting. To address this policy framework, several systems were implemented with the intention to provide an early warning of the onset of an epidemic and to prompt public health and military commander responses. However, the viability of some systems are now been questioned. For example, according to BioWatch documentation, 33.5 years of operational testing would be required to fully demonstrate that the system meets the established false positive rate.[6]

In the past, funding for biosurveillance was assigned to different government agencies, including the US Department of State (through the Agency for International Development), the Department of Defense (through the Defense Threat Reduction Agency, DTRA), the Medical Intelligence Center, the Global Emerging Infections Surveillance and Response System (DoD-GEIS), the Department of Human and Health Services (through the Centers for Disease Control and Prevention, CDC), the Department of Agriculture (through the National Animal Health Laboratory Network), and the Department of Homeland Security (through the National Biosurveillance Integration Center), and the US Geological Survey and the Biosurveillance Indications and Warning Analytic Community (BIWAC). In addition, international efforts, universities, and non-state and non-governmental organizations, such as the European Centre for Disease Prevention and Control (ECDC), Public Health Agency of Canada (PHAC), Panamerican Health Organization (PAHO), Bill and Melinda Gates Foundation, the World Bank, and Médecins sans Frontières, all contribute to event-driven biosurveillance efforts around the world.[7] The General Accounting Office's evaluation of different biosurveillance programs stressed the level of effort fragmentation and highlighted the need for strategic oversight mechanisms.[8,9] While a diversity of R&D biosurveillance programs have been discussed and implemented within the scientific community in the US and around the world,[10,11] an integrated technical strategy to address the policy requirements and systematically identify vulnerabilities across the entire government enterprise remains unfulfilled. As pointed out by the National Academy of Sciences, despite the recognition of its importance, the definitions and boundaries of biosurveillance activities vary in perspective by the stakeholders, who have different priorities and information needs, that might or might not reflect longer-term goals.[12,13] Despite the call for incorporating public health expertise into fusion centers to promote information sharing, the technical requirements for using open source information and its integration with human and zoonotic diseases reporting, remain uncertain.[14] After years of discussion, there is no consensus on preferred methodologies, performance characteristics, or outcome evaluation measures.[7,10] Furthermore, no formalized process currently exists among US government agencies and other organizations to coordinate efforts and facilitate collaboration among data generators, modelers, and decision makers.

The popularization of social media has fundamentally changed how individuals interact in our society. This has prompted several groups to develop data-mining tools to analyze sentiment, opinion and detect sudden outbreaks or shifts in infectious disease trends.[15-17] For example, the Personalized Tweet Ranking

Algorithm for Epidemic Intelligence (PTR4EI) provides users a personalized short list of tweets in the context of an infectious agent.[16] While impressive in numbers, real-time content generation from end-users hides much of the complexity in the factual reality. A large number of people might be discussing a natural phenomenon, but the semantic dimensionality of these interactions might not reflect the dynamics of the event itself. Quality issues within freeform user-provided hashtags and biased referencing can significantly undermine estimations of herding behavior, since the terminology to describe a disease by different people is a semantically volatile domain.[18,19] Persons engaged in social media tend to undervalue small probabilities and overvalue high probabilities.[20,21] Furthermore, some methods for extracting and integrating specific hashtags use the influence of trend persons to the determine the relevance of a particular subject or context.[22] Since topics attract users in an asymmetric way, word-of-mouth over social networks can be noisy and disproportionately disturb many of the algorithms mining this data.[21] These systems might not be as reliable when analyzing natural events and discriminating those from human actions and while they are a complement, they do not substitute traditional epidemiological surveillance networks.[23] Since it is difficult for the non-specialist to rapidly confirm the validity of each trend, the confidence in the information required to make critical decisions is significantly undermined. This situation is complicated by the fact that there are significant gaps to representing the analysis of this information in ways meaningful to aid decision-making. While it is clear that biosurveillance efforts must be integrated into the overall response system, few attempts have been made to rationalize this enterprise to combine social media data mining, geo-referencing, and molecular-based signal analysis. In this regard, this document presents an operational biosurveillance overview and discussions to prioritize new initiatives and existing investments that can generate well-informed tactical and strategic information to protect both the general population and warfighter. Emphasis is placed in the development of genomic-based decentralized medical intelligence and laboratory systems to track emerging and novel microbial health threats in both military and civilian settings and the use of virulence factors for risk assessment.

The Biothreat Landscape from a Medical Intelligence Perspective

Infectious diseases have played a significant role in the operational capability of armed forces, as the outcomes of war and combat illnesses continue to be affected by pathogens. As military commanders understood the implications of microbes, offensive bioweapon development programs emerged in more than 20 nations.[24] This proliferation of bioweapons, and their subsequent international prohibition, opened new offensive options for small groups engaged in irregular warfare against the dominance of US high-technology.[25] As biotechnology becomes increasingly de-skilled and less expensive, the proliferation of a new generation of biological weapons can now be easily achieved by state and non-state institutions, and even individuals.

Table 1. Biothreat technology assessment based on genomic metadata

	No. of country sources	Researchers	Institutions	Ratio
Bacillus anthracis	11	620	93	6.7
Orthobunyavirus	33	900	63	14.3
Ebola	7	318	16	19.9
Marburgvirus	5	185	7	26.4
Francisella tularensis	6	592	21	28.2
Flavivirus	139	6150	321	19.2
Tick-borne encephalitis	19	312	31	10.1
Orthopoxvirus	50	1420	261	5.4
Variola	28	107	5	21.4
Monkeypox	10	98	10	9.8
Monkeypox Zaire-96	1	14	4	3.5
Arenavirus	20	243	23	10.6
Coxiella burnetii	7	218	21	10.4
C. burnetii Dugway	1	22	4	22
Hantavirus	45	600	112	5.4

Therefore, the medical intelligence community needs to detect, assess, and foresee the status of technological development and the biothreat landscape in battlefield and civilian environments. However, design flaws of several components of the biosurveillance enterprise make it obvious that many operators are not grasping the long-range implications of dual use scientific and technological developments. For example, how will nanotechnology, information technology and their respective sub-fields alter the capacity to detect, track, and assess the risk of unknown and combinatorial pathogens.[2] While a variety of information gathering and signal processing techniques can be deployed for nuclear, radiological, and chemical threats, it is no secret that the collection of intelligence regarding biological weapon development and transference is one of the hardest tasks for analysts. This is because the "proliferation footprint of bioweapons" when compared with nuclear, radiological, and chemical threats is very small, and because, in short time periods, some countries can divert existing scientific expertise, experimental techniques and biotechnological facilities for offensive purposes (**Table 1**). Given the complexity of this situation, it could years for the intelligence community to understand the technological dynamics and complexity of the new biothreat landscape and even more years for the scientific groups to develop effective countermeasures against them.

Intelligence managers know that gaps exist within the information of a particular domain and from data derived from collectors and analysts, and that sometimes raw reports from human sources are fragmentary and biased, or just plain wrong.[26-29] Molecular-based assays and DNA sequencing should lead to portable high-resolution microbial typing methods that could be exploited for pathogen source tracing, attribution, and forensics. However, their impact in the battlefield will not be fully realized until standards ensure access to these signals by the warfighter.

Such a system must address the impracticality of transferring the terabytes of genomic data generated by each DNA sequencing device to a centralized architecture performing analysis operations, as that might take hours or even days. Therefore, a new paradigm could emerge from encouraging the development of decentralized algorithms that first determine in situ the presence of specific pathogen-specific genomic signatures or motif fingerprints, summarize and relay the results into an operational biosurveillance metadata format contextualized for military commanders and soldiers.

In a second stage, the biosurveillance metadata generated by different genomic-based analysis systems could be integrated and cross-validated by other near-sensors in a federated architecture using registries and semantic web technologies.[30] A key aspect for signal verification is the inclusion of pathogen-specific genomic signatures or motif fingerprints derived from the sample itself and directly associated with threat. At this level of resolution, a reference database of validated signatures of known pathogens isolated in natural events and from environmental sampling, sentinel organisms, as well laboratory culture and animal passage conditions can be used to determine the characteristics of the threat. This analysis process requires the correction and disambiguation of metadata associated with publicly available and confidential pathogen genomic sequencing efforts.[31-33] Since in many cases the information cannot be used as baseline for risk assessment, the analyst's support tools must include a wide range of software "plug-ins" integrating demographic distributions, pathogen characteristics, and availability of countermeasures into a low-probability-high-impact warfighter operational biosurveillance and forecasting framework. However, it is important to understand that different types of users perceive, synthesize, and use information in different ways to make decisions or influence decision making. Therefore, not knowing

the background logic between different attributes and the final prediction result might lead to the development of assessments lacking a solid support theory. To avoid one model that works well in one case and not in others, forecasting techniques must provide explicit probability statements of uncertainty. Unlike deterministic (single-value or yes/no), probabilistic modeling allow flexibility in the information content communicated to users based on their specific needs and preferences. Furthermore, the data analysis processes must operate in parallel rather than sequentially, and must be capable of reverse-engineering a new biothreat and quickly and efficiently propose target sites for countermeasure development. Forecasting and modeling products will be driven by a wide array of information requirements to be conveyed at different levels of sophistication. A new operational visualization formats should present highly technical information in a clear picture meaningful for formulating new policies and responses.

Biosurveillance through a Laboratory Network

Laboratory-based surveillance is pivotal to detecting and tracking infectious disease threats, since it relies on aggregating microbiological data at clinical care level and is supplemented by reference laboratory testing.[34] The ability to monitor the geographical spread of pathogens in cities, countries, continents or the globe can provide a perspective of the dissemination of a particular outbreak.[35] Considering the impact of public health decision making, the National Biodefense Science Board (NBSB) issued a series of recommendations including an oversight authority to assure compatibility, consistency, continuity, coordination, and integration of all the disparate systems associated with biosurveillance. The NBSB also recommended that the Secretary of HHS designate a central situational awareness authority for coordinating all the public health situational awareness data that has already been collected, processed, and analyzed from respective agencies on a national level. However, there is a need for clarifying what and how data regarding zoonotic, agricultural, and other potentially public health impacting events should be communicated and integrated into this idealized platform. For example, there are approximately 2300 hospitals and clinics and 160 reference laboratories monitoring infectious diseases in the United States. These entities have different preparedness levels to respond to acts of biological terrorism, emerging infectious diseases, and other public health emergencies, including the discrimination of known and unknown pathogens.[10] While some institutions use the Real-time Outbreak and Disease Surveillance Laboratory (RODS), Hospital Admission Syndromic Surveillance (HASS), Early Aberration Reporting System (EARS), Argus Biosurveillance System,[36] PulseNet,[37] Global Public Health Intelligence Network (GPHIN), HealthMap, and MedISys, these implementations do not operate at a level of sufficient resolution to be integrated with different FDA-approved and experimental pathogen detection devices and assays. Although diagnostic laboratories do submit strains or samples to reference laboratories for characterization and typing, many disease-causing pathogens of pandemic potential initially test negative with available testing panels. This is a key issue, since the world is confronted by new infectious agents that might pass undetected and circulate for a prolong time before they are recognized as such.

The development of high-throughput DNA sequencing technologies is allowing the genomic characterization of previously unknown pathogens without relying on prior reference molecular information.[13,38] This information is available within days, and even hours, of sample collection, and well before the development of animal infection models. Because of their portability, this technology will become widely used in the next 5 years in routine clinical settings. However, to be clinically and epidemiologically relevant in the biosurveillance context, DNA sequences must be rapidly and effectively translated into actionable information defining pathogen characteristics (i.e., virulence or drug resistance), it must point to a source of origin, and discriminate it from a natural event vs. a manmade release.[33] However, while some government agencies are considering use of genomic information to develop next generation Level 0 and Level 1 detection/surveillance devices,[6] there is no reference database where researchers can retrieve standardized genomic signatures and motif fingerprints to develop primer-, probe-, and antibody-based detection technology. While such information will directly impact threat level assessment and the prioritization of medical countermeasures, the use of genomic signatures or their corresponding amino acid motif fingerprints could lead to standardized and interoperable detection technologies.[2] For example a motif fingerprint schema allows a bioforensic-attribution system for inclusion or exclusion pathogens based on binary patterns specific to a taxonomic level. Thanks to this approach it is possible to narrow individual genomic signatures associated with virulence, transmission mode, reservoirs, and hosts to assess threat risk level (**Fig. 1**).

Developing metadata standards for an operational biosurveillance requires addressing discrepancies in the taxonomic assignments between public databases. On average, 11% of records used for reagent/assay development and pathogen identification schemes are incorrectly assigned within a particular taxonomy (i.e., serotype or species).[32] The current system of nomenclature used for the classification of viruses is a significant limitation to understanding the evolutionary history of many such pathogens; therefore, a modified taxonomic numerical system could facilitate the tracking of new pathogens in a biosurveillance environment.[32,39]

Since only certain diseases caused by infectious agents must be reported, the identification of a new pathogen requires optimizing the risk thresholds for select agent categorization as data of these pathogens becomes available. While international surveillance networks rely on reference laboratories, each pathogen or pathogen group has its own network and analysis system, often with a centralized data collection system that follows unique standards. As more and more clinical laboratories perform molecular testing using next generation sequencing
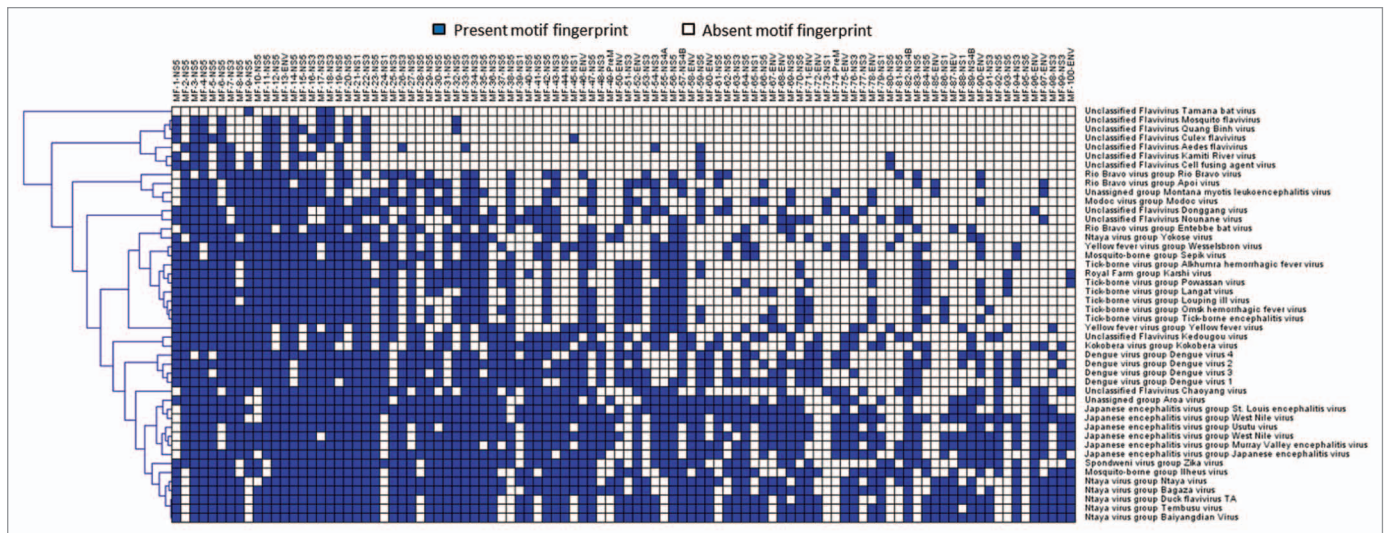
Figure 1. Motif fingerprinting of viruses. The *Flavivirus* genus comprises species responsible for several emerging and re-emerging diseases. The short replication times and high mutation rates of these viruses have hampered attempts to isolate genome segments that can be associated with their origin, form of transmission, and pathogenesis. A computational survey of all available sequence information for this genus identified species-specific protein motifs fingerprints. The presence of these genomic elements forms binary patterns that provide a new framework for taxonomical classification.

technologies, the reference laboratories become dependent on data submission. This situation is blurring the distinction between diagnostic and reference/public health laboratory functions and challenges the hierarchical architecture of reference laboratories, since, at the international level, few benefits are obtained by the submitting laboratories on the translational impact of their samples.

A largely unresolved question is how genome sequences must be examined for epidemiological characterization.[40] Bioinformatics and computational biology advances of the last two decades have led to an increase in the number of databases for microbial typing.[35] However, the narrow number of microbial species supported in the database, lack of interoperability, and the proprietary schema of many of these efforts requires new formats that federate this information within a biosurveillance enterprise.[41,42] The development of most data management systems focuses on the current state of the technology without considering how their design will affect the legacy of sensors and assays. At the moment, there is no single ideal pathogen discrimination and genotyping approach, nor standards and benchmarks, available under national and international settings.[2,35,40,43] Optimizing a system of threshold detection-based sensors, in the sense of maximizing the probability of detecting an event of interest, is subject to a constraint on the expected number of system-wide false signals. Existing tools used for genomic analysis of metagenomic samples are largely unsuitable for biosurveillance: (1) they suffer from high false positive or false negative error rates (ranging from 15% to 80%), (2) even the most sensitive analytical tools fail to identify 20% of test data sets, and (3) with existing algorithms, the analysis of next generation sequencing data can take several days.[2] Therefore, microbial databases must evolve as metadata-compatible biosurveillance systems with translational support that

not only characterize outbreaks and trace evolutionary pathways but also guide countermeasure development (**Fig. 2**). This architecture must address the open nature of data submission and the different degrees of reliability of different diagnostic assays and algorithms. While cloud computing presents an obvious framework to address current demands of storing and processing big data, "genomic data streams" need to be addressed, not only by increases in hardware, but by encouraging the development of new and efficient algorithms capable to operate in situ and a federated fashion.

Conclusions

A critical function of operational biosurveillance is the ability to rapidly, reliably, and securely collect, synthesize, and share diverse sources of information among medical intelligence, public health, military commanders, and decision makers. While it is acknowledged that biosurveillance can provide a comprehensive picture of the health status of military personnel deployed in a particular area, as well as in national and international communities, the integration of clinical and genomic information across multiple levels of government, professional practices, and scientific disciplines represent a significant challenge. Nonetheless, a genomic-based biosurveillance awareness system integrating routine microbial genotyping for virulence holds the potential to accelerate recognition of a pathogen's virulence, enable a rapid, targeted intervention, and guide the development of additional countermeasures. The extent to which isolates can be compared depends not only on the quality of the sequence data available, but in the quality of the "data about the data" and the implementation of decentralized analysis systems. Considering the significant amount of data already being generated at the
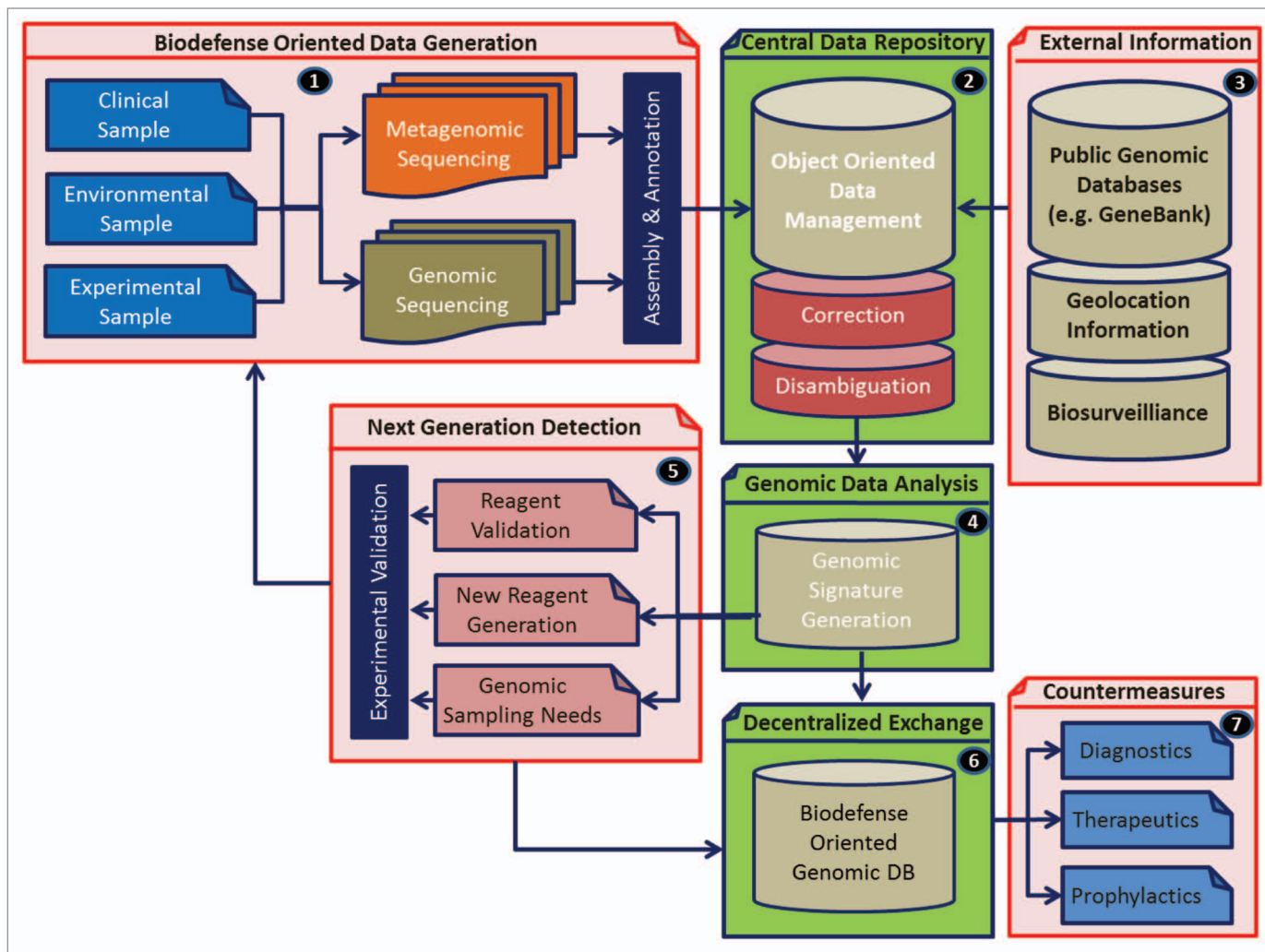
Figure 2. Biodefense enterprise system for global pathogen awareness and countermeasure deployment. This global awareness system will integrate information collected by software agents and artificial intelligence algorithms capable of prioritizing and classifying pathogen genomic information and its associated metadata to yield specifics of potential actors, their capabilities, and potential feasibilities. This effort consists in the integration, annotation, disambiguation, evaluation, and representation of genomic and open source metadata information to conduct assessment including available and projected capability. This differs from other biosurveillance techniques that are assembled (and evaluated) for the purpose of pathogen detection and prediction without considering technology and future trends of technological capabilities. The system includes the most likely and most stressing threats and identify intelligence gaps (if any) that can affect the efficacy of any countermeasure program.

metagenomic level by next generation DNA sequencing, a centralized analysis system receiving terabytes of data streams is impractical. Therefore, more attention should be placed in developing a decentralized analytic system, with capabilities to rapidly discriminate anomalous profiles of genomic information. This operational biosurveillance capability must support data integration, ubiquitous metadata sharing, communication networking, advanced analytics, and data representation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. DiGiovanni C Jr. Domestic terrorism with chemical or biological agents: psychiatric aspects. Am J Psychiatry 1999; 156:1500-5; PMID:10518158

2. Valdivia-Granda WA. Bioinformatics for biodefense: challenges and opportunities. Biosecur Bioterror 2010; 8:69-77; PMID:20230234; http://dx.doi.org/10.1089/bsp.2009.0024

3. National Research Council (U.S.). Committee on Advances in Technology and the Prevention of Their Application to Next Generation Biowarfare Threats., Institute of Medicine (U.S.). Board on Global Health. Globalization, biosecurity, and the future of the life sciences. Washington, D.C.: National Academies Press, 2006.

4. Simon JD. Rand Corporation., U.S. Armed Forces Medical Intelligence Center. Terrorists and the potential use of biological weapons: a discussion of possibilities. Santa Monica, CA: Rand Corporation, 1989.

5. Vogel KM. Phantom menace or looming danger?: a new framework for assessing bioweapons threats. Baltimore: Johns Hopkins University Press, 2013.

6. Jenkins WO. United States. Congress. House. Committee on Homeland Security. Subcommittee on Emergency Preparedness Response and Communications., United States. Congress. House. Committee on Homeland Security. Subcommittee on Cybersecurity Infrastructure Protection and Security Technologies., United States. Government Accountability Office. Biosurveillance observations on BioWatch Generation-3 and other federal efforts: testimony before the Subcommittees on Emergency Preparedness, Response, and Communications and Cybersecurity, Infrastructure Protection, and Security Technologies, Committee on House Homeland Security, House of Representatives. Testimony GAO-12-994T. Washington, D.C.: U.S. Govt. Accountability Office, 2012.

7. Hitchcock P, Chamberlain A, Van Wagoner M, Inglesby TV, O'Toole T. Challenges to global surveillance and response to infectious disease outbreaks of international importance. Biosecur Bioterror 2007; 5:206-27; PMID:17903090; http://dx.doi.org/10.1089/bsp.2007.0041

8. United States. Government Accountability Office. Biosurveillance nonfederal capabilities should be considered in creating a national biosurveillance strategy: report to congressional committees. Washington, D.C.: U.S. Govt. Accountability Office, 2011.

9. United States. Government Accountability Office. Biosurveillance DHS should reevaluate mission need and alternatives before proceeding with BioWatch Generation-3 acquisition: report to congressional requesters. Washington, D.C.: U.S. Govt. Accountability Office, 2012.

10. Kman NE, Bachmann DJ. Biosurveillance: a review and update. Adv Prev Med 2012; 2012:301408; PMID:22242207; http://dx.doi.org/10.1155/2012/301408

11. Zeng D, National Science Foundation. (U.S.). Intelligence and security informatics–biosurveillance: second NSF workshop, BioSurveillance 2007, New Brunswick, NJ, USA, May 22, 2007; proceedings. Berlin: Springer, 2007.

12. Patel D, Olson S, Institute of Medicine. (U.S.). Planning Committee on Information-Sharing Models and Guidelines for Collaboration: Applications to an Integrated One Health Biosurveillance Strategy., Institute of Medicine (U.S.). Board on Health Sciences Policy. Information sharing and collaboration applications to integrated biosurveillance: workshop summary. Washington, D.C.: National Academies Press, 2012:1 online resource (xii, 99 p.).

13. Ronald D, Fricker J, Banschbach D. Optimizing biosurveillance systems that use threshold-based event detection methods. Inf Fusion 2012; 13:117-28; http://dx.doi.org/10.1016/j.inffus.2009.12.002

14. United States. Congress. House. Committee on Homeland Security. Subcommittee on the Prevention of Nuclear and Biological Attack. Creating a nationwide integrated biosurveillance network: hearing before the Subcommittee on Prevention of Nuclear and Biological Attack of the Committee on Homeland Security, House of Representatives, One Hundred Ninth Congress, second session, May 11, 2006. Washington: U.S. G.P.O.: For sale by the Supt. of Docs., U.S. G.P.O., 2007.

15. Diaz-Aviles E. Avar, #233, Stewart. Tracking Twitter for epidemic intelligence: case study: EHEC/HUS outbreak in Germany, 2011. Proceedings of the 3rd Annual ACM Web Science Conference. Evanston, Illinois: ACM, 2012:82-5.

16. Stewart A, Diaz E. Epidemic intelligence: for the crowd, by the crowd (Tutorial), in 'Proceedings of the 12th international conference on Web Engineering', Springer-Verlag, Berlin, Heidelberg, pp. 504--505.

17. Liu B. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, San Rafael, Calif. (1537 Fourth Street, San Rafael, CA 94901 USA): Morgan & Claypool, 2012:1 online resource (xiv, 167 p.).

18. Sahlgren M, Karlgren J. Terminology mining in social media. Proceedings of the 18th ACM conference on Information and knowledge management. Hong Kong, China: ACM, 2009:405-14.

19. Gayo-Avello D. Don't turn social media into another 'Literary Digest' poll. Commun ACM 2011; 54:121-8; http://dx.doi.org/10.1145/2001269.2001297

20. Wong FMF, Sen S, Chiang M. Why watching movie tweets won't tell the whole story? Proceedings of the 2012 ACM workshop on Workshop on online social networks. Helsinki, Finland: ACM, 2012:61-6.

21. Kim YA, Ahmad MA. Trust, distrust and lack of confidence of users in online social media-sharing communities. Knowl Base Syst 2013; 37:438-50; http://dx.doi.org/10.1016/j.knosys.2012.09.002

22. Potts L, Seitzinger J, Jones D, Harrison A. Tweeting disaster: hashtag constructions and collisions. Proceedings of the 29th ACM international conference on Design of communication. Pisa, Italy: ACM, 2011:235-40.

23. Butler D. When Google got flu wrong. Nature 2013; 494:155-6; PMID:23407515; http://dx.doi.org/10.1038/494155a

24. Wendl MC. Bioweapons could kill more in one strike than guns. Nature 2005; 437:192; PMID:16148911; http://dx.doi.org/10.1038/437192b

25. Ouagrham-Gormley SB, Vogel KM. The social context shaping bioweapons (non)proliferation. Biosecur Bioterror 2010; 8:9-24; PMID:20230229; http://dx.doi.org/10.1089/bsp.2009.0054

26. Walsh JI. The international politics of intelligence sharing. New York: Columbia University Press, 2010.

27. Harvey FP. Explaining the Iraq War: counterfactual theory, logic and evidence. Cambridge; New York: Cambridge University Press, 2012.

28. Jervis R. Why intelligence fails: lessons from the Iranian Revolution and the Iraq War. Ithaca: Cornell University Press, 2010.

29. United States. Central Intelligence Agency. Comprehensive report of the Special Advisor to the DCI on Iraq's WMD. Washington, D.C.: Central Intelligence Agency, 2004.

30. Sinaci AA, Laleci Erturkmen GB. A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains. J Biomed Inform 2013; 46:784-94; PMID:23751263; http://dx.doi.org/10.1016/j.jbi.2013.05.009

31. Valdivia-Granda W. The next meta-challenge for Bioinformatics. Bioinformation 2008; 2:358-62; PMID:18685725; http://dx.doi.org/10.6026/97320630002358

32. Valdivia-Granda W, Larson F. ORION-VIRCAT: a tool for mapping ICTV and NCBI taxonomies. Database (Oxford) 2009; 2009:bap014; PMID:20157487; http://dx.doi.org/10.1093/database/bap014

33. Valdivia-Granda W. Biodefense Oriented Genomic-Based Pathogen Classification Systems: Challenges and Opportunities. J Bioterr Biodef 2012; 3:2-9

34. Struelens MJ, Brisse S. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill 2013; 18:20386; PMID:23369387

35. Carriço JA, Sabat AJ, Friedrich AW, Ramirez M; ESCMID Study Group for Epidemiological Markers (ESGEM). Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. Euro Surveill 2013; 18:20382; PMID:23369390

36. Nelson NP, Yang L, Reilly AR, Hardin JE, Hartley DM. Event-based internet biosurveillance: relation to epidemiological observation. Emerg Themes Epidemiol 2012; 9:4; PMID:22709988; http://dx.doi.org/10.1186/1742-7622-9-4

37. Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyytiä-Trees E, Ribot EM, Swaminathan B; Pulsenet Taskforce. PulseNet USA: a five-year update. Foodborne Pathog Dis 2006; 3:9-19; PMID:16602975; http://dx.doi.org/10.1089/fpd.2006.3.9

38. Gallego B, Sintchenko V, Wang Q, Hiley L, Gilbert GL, Coiera E. Biosurveillance of emerging biothreats using scalable genotype clustering. J Biomed Inform 2009; 42:66-73; PMID:18723122; http://dx.doi.org/10.1016/j.jbi.2008.07.002

39. Firth C, Tokarz R, Simith DB, Nunes MR, Bhat M, Rosa ES, Medeiros DB, Palacios G, Vasconcelos PF, Lipkin WI. Diversity and distribution of hantaviruses in South America. J Virol 2012; 86:13756-66; PMID:23055565; http://dx.doi.org/10.1128/JVI.02341-12

40. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl Jm, Laurent F, Grundmann H, Friedrich AW; ESCMID Study Group of Epidemiological Markers (ESGEM). Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 2013; 18:20380; PMID:23369389

41. Callahan A, Cruz-Toledo J, Dumontier M. Ontology-Based Querying with Bio2RDF's Linked Open Data. J Biomed Semantics 2013; 4(Suppl 1):S1; PMID:23735196; http://dx.doi.org/10.1186/2041-1480-4-S1-S1

42. Seoane JA, Aguiar-Pulido V, Munteanu CR, Rivero D, Rabunal JR, Dorado J, Pazos A. Biomedical data integration in computational drug design and bioinformatics. Curr Comput Aided Drug Des 2013; 9:108-17; PMID:23294434

43. Bohlin J. Genomic signatures in microbes -- properties and applications. ScientificWorldJournal 2011; 11:715-25; PMID:21442149; http://dx.doi.org/10.1100/tsw.2011.70