Emerging Methods in Public Health Informatics:

Addressing the Collaboration Gap

Susanne Jul, PhD, Robert Kirkpatrick, Judith Kleinberg, Dennis Israelski, MD and Eric Rasmussen, MD, MDM, FACP

InSTEDD

Abstract

In Cambodia, a poor country in a region vulnerable to potential pandemic-producing outbreaks (H5N1, SARS), baseline health surveillance circa 2008 begins with village health workers collecting data from local village elders. Next, the health workers travel by bicycle or boat to a district clinic to transcribe the reports, which are then sent to the provincial medical office via high-frequency radio where they are entered into WHO's EWARN database in English by a clerk who barely knows the language. The information is then fed by EWARN to the country's health ministry using modems on dial up land line phones over a computer system powered part of the time by car batteries.

The system is technically fragile, prone to transcription errors, time and labor intensive, requires knowledge of at least two languages (Khmer and English), and offers little opportunity for feedback.

We can do better.

InSTEDD's work is currently focused in Cambodia, but the technologies and system architecture described in this white paper for developing a more effective approach to health surveillance in a country with extremely limited resources can be adapted almost anywhere. Indeed, only when surveillance systems are technologically harmonized to allow for timely data analysis will the goal of predicting and preventing global public health crises be realized.

Public Health Informatics (PHI) is an umbrella term for technologies and methods to improve the sensitivity, specificity and speed at which health threats are identified. There are four main components to PHI: discovery, analysis, decision-making and collection action.

It is a collaborative effort. Success depends just as much on socio-cultural factors as it does on technologies. Social structures that support disease reporting are the bedrock of the system. Since surveillance data can have politically sensitive economic as well as health implications, care must be taken to create a system with safeguards against false alarms. The threat may be global – or regional – but the initial costs, whether of lives or livelihoods, are local. Likewise, a balance between transparency, accountability and data security is critical for securing governmental "buy-in" to a PHI design.

This paper begins with a description of 10 problems areas in biosurveillance. We then describe a grid of five primary information domains that help with those problems: Information flow, Mesh synchronization, Analysis, Decision Support and Collaboration. The five domains are each subdivided into roughly six sub-domains. For example, "Social Networking" and "Messaging" are part of "Collaboration," while "Translation," "Deep Field Collection" and "Geocoding" are part of "Information Flow."

There is, naturally, considerable overlap between domains. Being better able to visualize the whole, however, can make it easier to see what is needed where and to develop strategies to make it happen.

Making it happen is primary interest for InSTEDD. We pay special attention to the use of technology for outbreak detection and response improvement for the disease hotspots of the developing world. We perceive an unmet need for capabilities that facilitate a self-organizing transition from event detection, to collaborative sensemaking, to collective action and we think that resources need to be put toward that goal.

Our further opinion is that systems for PHI in the developing world are best adopted by practitioners when designed for ruggedized resilience, a spectrum of users, and for dual-use. The introduction of

dual-use technologies, and social structures for effective outbreak reporting, using methods that have persistent value even when not being used for that explicit purpose. Communications put in place to meet the International Health Regulations (IHR) reporting requirements could also serve as a starting point for improved governance, encouraging greater transparency and accountability and perhaps driving expanded regional cooperation.

We have not overlooked the fact that the capabilities we describe below for public health informatics have similar value in the reporting of *indirect* health issues. Environmental health threats from slash-and-burn clearance smoke; freshwater contamination from sewage leaks, mine tailings and oil field seepage; fish losses from agricultural and pharmaceutical toxins; acute and chronic illness from industrial disasters like Bhopal and Chernobyl, all could be more effectively managed with the information tools we advocate.

We will discuss below our view of the challenges and opportunities we see in the technical and socio-cultural aspects of PHI for outbreaks. We will also describe, with some discussion, the core problem set and information-processing model we find useful for evaluations within rural Southeast Asia, a region of excess risk for emerging and resurgent infectious diseases and the focus of much global health security concern.

Key Words

Public Health Informatics, biosurveillance, collaboration, synchronization, global health, algorithm, emerging infectious diseases, security, technology, cellular technology, satellite technology, resilience, collaboration gap

Introduction to the collaboration problem

International public health security, a subject of the World Health Organization Annual Report in 2007, can be described as a collective aspiration and a mutual responsibility. This goal and our mutual responsibilities require a capacity for shared discovery, analysis, and decision-support leading to collective action.

These capacities require a collaborative framework that, as far as we can ascertain, does not currently exist at any district, province, state, national or international level. “Collaborative framework” means the rapid and seamless incorporation of the best thinking on urgent information by a range of experts and interested parties, leading to effective collective action that mitigates the severity of an infectious disease outbreak. When global trade and travel can place any pathogenic microbe in any national capital within 20 hours, bridging that collaboration gap is critical.

We have a staff experienced in fieldwork associated with biosurveillance and the medical aspects of informatics in disaster response. With that experience we’ve compiled a list of ten common problems that have arisen recurrently, seeming to need deeper consideration and perhaps academic research.

Platform development

More strategically, and with international public health security as our goal, we have attempted a distillation of the essential capabilities required for outbreak detection and response within the modern world. We’re working, with many partners, on how to integrate each of those pieces into a seamless whole; a coherent platform. A platform designed to include those elements, seen in the image below and described in the Appendix, could provide an extraordinary advantage in our capability to detect, assess, and respond to emerging and re-emerging disease threats.

The techniques and technologies available for such a collaborative framework have not yet coalesced into a comprehensive system. Based on conversations around the world, though, the convergence of the requisite disciplines is well underway.

Although we’re focused heavily toward technology in this discussion, it’s important to remember that the political and social barriers are often far greater than the technical barriers in most of the cases

we've examined, therefore incorporating methods for socio-cultural integration appears to be at least as valuable as developing the technical tools.

Layered requirements

Starting with the technical side, introducing new and potentially disruptive technical tools for risk assessment and mitigation can be done but requires integration into the policy, culture, and organizational processes of the appropriate sites. Information sharing may not be a natural impulse, but failures to share information well, especially where the impact of technology was misunderstood, have very likely led to excess mortality in events like Chernobyl, SARS, and Bhopal.

We recognize there is a tension in public health between the need for rapid and effective information flow between all parties with a need to know, and the confidentiality requirements of what is often sensitive medical information with a potential impact on individuals and economies, but options have been designed to manage it. Inculcating a culture of collaboration should be considered a goal at every stage of public health information flow. We will describe below a few useful tools and methods for encouraging that culture.

It is not possible to separate the design of information collection from the infrastructure requirements of electron flow (both power and communications). Nor should we overlook the implementation and response capability that policy and law provide. These three tiers, infrastructure, sense making, and policy, are interdependent and necessary at every stage of PHI in the developing world.

Once the core infrastructure is secured and policy support is expected, a next step in ensuring sense making in PHI is the capacity to communicate effectively across geographic, social, linguistic, cultural, religious and organizational boundaries, and over a range of devices and channels, even if the information flow is entirely within a single community, state, or nation. There are, of course, wide swaths of the world where sophisticated communication is only intermittently possible, even when the organizational will to do so is present. Communication structures that are designed to be flexible, modular, multimodal, and multilingual can help the public health system transcend such boundaries.

Leapfrogging

Over time those populations currently disconnected, both from the power grid and from the computational cloud, will gain connectivity. When they do, leapfrog technologies will likely be introduced using market-based forces that avoid the capital-intensive "copper wire" phase of development and take advantage of a demand for non-specific communications. Leapfrog adoption has expanded cellular telephone technologies, for example, to 3.3 billion phone subscriptions in 2007, 68% of them in the developing world. As of 2007, roughly 80% of the world's population was living within cellular range. That ubiquitous form factor cannot be ignored. We should be designing for cellular as the first technology at the farthest-forward boundary of reporting with tools that bridge from cellular to other data transmission modes seamlessly and transparently.

For real usefulness, any modern public health informatics tool should be familiar, inexpensive, accessible, robust and resilient. They should take advantage of dual-use opportunities so that the system remains familiar through daily use. It is also advantageous if the information tools are culturally and religiously sensitive, politically neutral, driven toward voluntary adoption through indigenous community structures, and offer opportunities for micro-economic development, distribution, maintenance and expansion.

Social networking tools are commonplace now, and the science of mesh communication has become quite robust. We have moved beyond monotonic information streams to complex rivers that interweave and loop back in ways that require collaborative assessments across a resource mesh. Designing for that mesh-like model will allow us to move beyond the limitations of ordinary reporting systems. We have seen the painful results of unanticipated information storms that overwhelm conventional hierarchical systems [Golden Shadow Report, InSTEDD, 2007]. Parallel streams of information in an emergency will soon become common and should be anticipated.

We need it all

We should also recognize the false tensions between personal healthcare documentation and population threat management. To get the best yield for the detection and mitigation of emerging threats, PHI needs an integrated and hybrid continuum of epidemiological surveillance across human personal medical records, human public health, pet and production animal health, and environmental health that includes wildlife. That synthesis is a tough problem, but not impossible, and the consequences of doing less might be severe.

Recognized problems in disease reporting:

Before we look at the 31 elements we've mentioned as useful, lets start with the problems we've seen.

The past ten years have revealed a consistent set of field-based problems associated with the collecting and reporting of disease outbreaks. While this list is idiosyncratic and undoubtedly incomplete, the InSTEDD staff agree on the few described here. The problem set for disaster response is similar. Teams across the globe are trying to address each of them.

1. Cultural acceptance
2. Geo-referenced imagery
3. Languages and machine translation
4. Unreliable communications
5. Minimal essential data sets
6. Complex Adaptive Systems modeling
7. Epidemiology decision support
8. Rapid assessment consolidation
9. Emergent strategic collaboration
10. Consolidating One Health impact

A few thoughts on activity within each…

1. Cultural acceptance

    a. We have found occasions where a cultural lifetime of implicit and explicit messaging has made it impossible to implement seemingly sensible and productive ideas in outbreak surveillance and disaster relief. There are recent examples in Myanmar, Afghanistan, and Iraq (with a particularly striking example from the Yogjakarta, Indonesia earthquake in 2006 when "White Monkey" stories were used to counteract gender discrimination that was leading to excess mortality) but many field workers would have similar stories from elsewhere. Having access to a cultural anthropologist or an ethno-sociologist can be very helpful in response planning.

2. Geo-referenced imagery

    a. Many organizations are trying to provide high-resolution imagery quickly to outbreak and disaster responders globally. Efforts are underway through the UN, national government, and nonprofit sectors, including UNOSAT, NASA Ames, NGA, and Google.org. There are also efforts through academic institutions like San Diego State University and Arizona State's Biodesign Institute. The private sector is contributing through services like Google Earth, Ikonos, Digital Globe, TerraServer, and others. The current goal at NASA, for example, is to provide recent imagery for public health and humanitarian support within 4 hours of the stated need, and to update with current (post-event) imagery within 48 hours for comparison overlays.

3. Languages and machine translation

    a. Many outbreaks occur in areas where Mandarin, English, Hindi, Arabic, and Spanish, the most common languages in the world, are not spoken. Recent language needs have included Burmese in Myanmar, Khmer in Cambodia, Yi and Kham in Sichuan, and 22 languages in Ecuador. Most outbreak responders with international agencies work only in English or French so communication impediments are severe, particularly in ground-based public health assessment and reporting.

    b. There are organizations working on machine-based translation, and the state of the art requirement as of June 2008 is apparently about 10 million words in parallel text (original language-accurate English). That seems to be enough to build a machine-based translation model that has usable accuracy. Actually, that's not much of a burden and is a reflection of improved linguistic algorithms and the processing power of portable computers. DARPA, Johns Hopkins University, Carnegie-Mellon, MITRE Corporation, and USC-ISI are a few of the sites leading research efforts in machine translation in the US. MITRE has a near-real-time multilingual chat client for medical interviews in 17 languages in use today. Google is expanding its web-based translation engine frequently, allowing anyone with web access to translate text with reasonable accuracy at no charge. The Google *Translate* Application Programming Interfaces (APIs) are also now open and available for adaptation by anyone.

    c. We would note that communication in machine-based translation is far better when using interactive modes, like chat clients, where both participants can inquire about awkward phrases immediately. It is also true that the inherent inaccuracies of machine translation lead to far more utility in subjects like public health (where getting the gist of the idea leads to filters that improve the efficient use of human translators) rather than in, for example, conflict mediation.

4. Unreliable communications

    a. Communications is a core competency in public health informatics. We recommend multiple communications modes, most of which should be in daily use for routine communications and known to everyone. The metaphor we use is that of the fabled "Red Telephone" in the bottom drawer, pulled out in the event of a disaster when the regular black telephone no longer functions. The problem with the Red Telephone is that it has metaphorical dials and switches and codes that are unfamiliar and so may not be remembered in an emergency. We believe that technologies that are in use daily, yet resilient and able to perform in crises, should be adopted as the standard.

    b. In most cases the bedrock foundation is SMS text messaging over cellular systems, as it was for the first week in Banda Aceh after the tsunami. The Kobe earthquake recommendations also emphasized the value of multimodal communications starting with SMS. We also recommend provisioning field staff with SMS tools that bridge to Internet resources seamlessly and to test the links frequently with ordinary messages. Microsoft Research India has developed such an SMS gateway and released it as open source for free.

5. Minimal essential data sets

    a. The data elements collected in field-based health assessments are often ad hoc, occasionally excessive, and frequently redundant. Such inefficient data collection is burdensome to those providing care and looks clumsy to those assessing impact. As International Strategy for Disaster Reduction and the Health Metrics Network argue, the lack of standardized collection methodologies and the lack of commonly accepted definitions together inject waste and confusion into just those areas where we need greatest clarity. Better options are under development for trials late this year, with continuing discussions at the HISA Durban Conference in June 2008.

6. Complex Adaptive System modeling

    a. Trans-generational complex emergencies arise from a multitude of factors. Example: Bangladesh is a very flat, crowded, poor country with 150 million people, and with more than 5000 square miles of the country lower than three feet above sea level. It has sea levels now rising from climate change, Himalayan freshwater glaciers melting too fast from particulate deposition after fires in Indonesia (so reducing fresh water access), blue-ear disease killing pigs, H5N1 decimating chickens (so both meat and eggs lost, and so most protein), rice supplements from WFP lost in the Myanmar cyclone, food prices doubling in the past 12 months, and frequent cyclones that contaminate tens of thousands of coastal freshwater sources with salt water, forcing internal migration.

    b. Complex emergencies are likely to increase around the globe and the health implications are obviously severe. PHI might benefit from participating in the ongoing convergence of climatology, sociology, anthropology, development economics and mathematics than we have seen thus far. We would also likely benefit from deeper relationships with academic disaster specialists like CRED in Belgium and ISCRAM in the Netherlands. The collaborative methods available for each specialty are becoming more competent. Mesh-based evaluations should be available, as a technique, later this year.

7. Epidemiology decision support

    a. The determinations of significance in epidemiology are crucial in deciding how to respond to an anomaly. Remember that, in statistics, **confidence** (what we're striving for…) is the result of solving the equation:

        [**(signal/noise) times the square root of the sample size**] = confidence level

        And that helps practitioners determine where to put system development efforts. This is nothing specific to the epidemiology of outbreaks – it's just clinical statistics 101, but a useful reminder of how to focus.

    b. A high level of confidence (a narrow confidence interval) implies the capacity to filter (optimizing the first term, as previously discussed) and having a large baseline sample from which to select for anomalies. The root of that sample size, though, changes slowly. 100 samples=10, 500 samples~22, 1000 samples~31, so clearly optimizing the SNR is the way to go in disease surveillance.

    c. Since we're talking about the need to reduce noise in the system so that we might save lives, such decision support begs for the best assessment possible. There are teams now working on hybrid models for combining case-based and indicator-based biosurveillance algorithms to improve both sample size and the management of the ratio. Initial efforts have been run against the Reuter's Reference Database with results that seem encouraging. The methods will be released as open source once validated to a first approximation, asking then for the community of practice to join in refining the techniques.

8. Rapid Assessment consolidation

    a. Minimal essential data set selection is an issue in virtually every assessment and response. There are now methods designed for software-based forms consolidation in Rapid Assessments. Although we know of no implementation trials in progress (there was a brief study in Darfur in 2006), the pursuit of such technologies and trials should be a listed priority in the research arm of public health informatics.

9. Emergent strategic collaboration

    a. Collaboration should be a core competency in public health. There are multiple methods for achieving a reasonable degree of shared understanding quickly, both socially and

technically. Dozens of tools are available as both commercial and open source software. A partial list of collaborative resources would include Groove, Ning, Riff, Google Groups, Trac, Basecamp, Adobe Connect, Central Desktop, Elluminate, Open Teams, Zing, and WebEx. A limitation of most of these systems is their inability to integrate with health surveillance systems, open media analyses, sensor networks, or event management tools. And only a very few are designed to facilitate collaborative sense making across sectors.

b. At InSTEDD we use Groove internally, Google Groups externally, Ning for demonstrations, and Riff (an internal InSTEDD product), depending on the task.

10. Consolidating across human-animal-environmental health

a. This has been addressed above, but we'll note that SARS was spotted in the veterinary literature months before the first human case appeared. The same was true in the reporting of Nipah in 1999, and in the 1996 outbreak of H5N1 avian influenza in Chinese geese. Better collaboration across the human-animal divide could lead to earlier detection and a more rapid response to biological threats globally.

Technical Opportunities and Challenges in PHI: biosurveillance and response:

Let us briefly discuss the 31 technical elements in the five domains that were mentioned above. A matrix of the elements can be found in the image below.

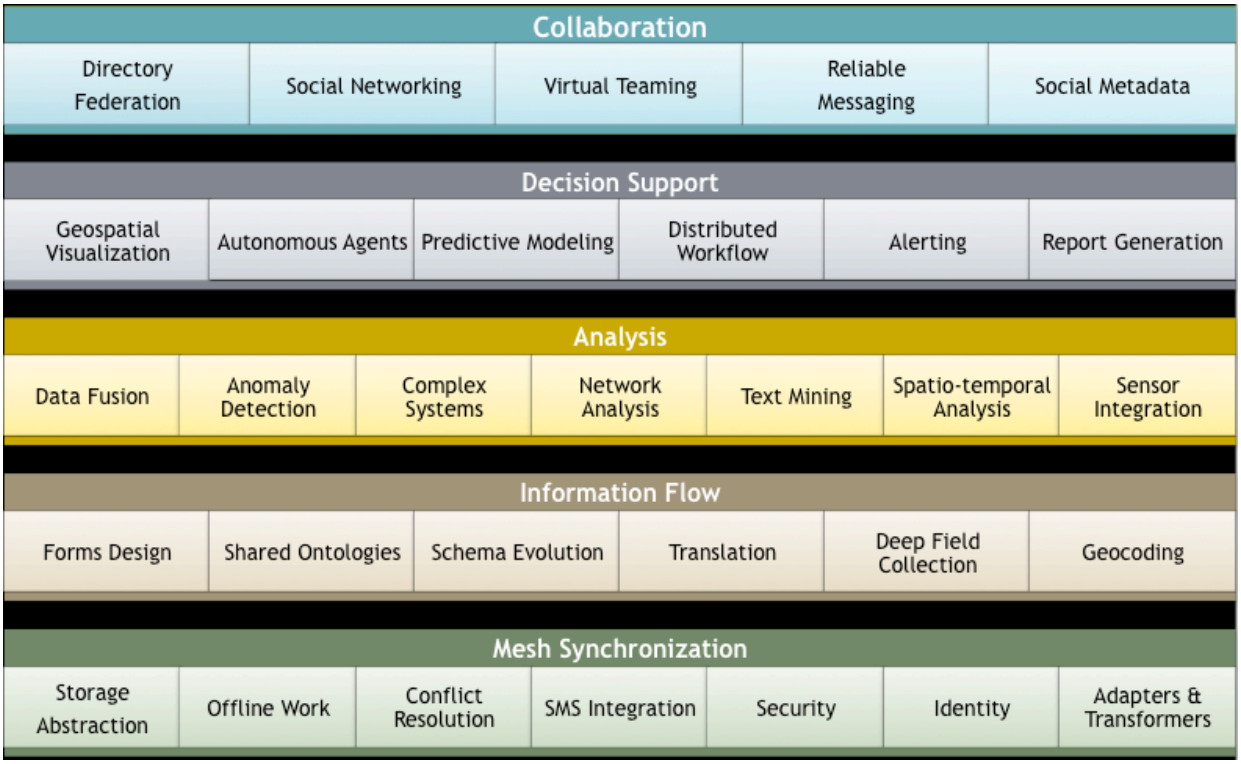
We consider these items the reference framework needed for a modern PHI system designed to optimize information flow in biosurveillance, mitigating the impact of outbreaks. In our opinion, this system needs to be built.

As seen in the image, the five information domains for those elements are:

- Information flow, mesh synchronization, analysis, decision support, and collaboration

The matrix below shows each information domain building upon the other. Each of those domains requires its own characteristic set of elements, and our research has shown that all of the items depicted have either some research effort in progress or the item is already available for use in the field. Since we recognize that many of the element terms are unfamiliar, there is a description of each one in Appendix 1.

Again, the image below and the descriptions in the Appendix are not intended as an arcane glossary of obscure techniques in information science. We intend these as a productive list of attributes for tools applied toward outbreak detection, enhanced analysis, and the promotion of effective collective action against communicable diseases of international concern.



Implicit in every stage of development for this matrix is the presence of **metrics** to ensure we're really accomplishing what we intended, with feedback loops and iteration. A PHI platform designed on this matrix is an achievable goal in the foreseeable future.

Stages:

Practitioners in PHI recognize the stages of data collection, data transmission, noise reduction, education and insight, decision, judgment, and collective action. To examine the opportunities and challenges within that set we can use the following scenario:

InSTEDD is, genuinely, working in northern Cambodia. We work with clinics supporting remote Community Village Health Workers who go out from the district health centers to collect health surveillance information from local village elders. The Health Workers, on foot, on bicycles or in small rowboats, then return to the district clinic to transcribe a verbal report. That report then goes out over HF radio to the Provincial Medical Office that has a clerk enter it into WHO's EWARN software (in English, a language in which the clerk has minimal proficiency). EWARN then feeds the information through local dial-up land lines to the Ministry of Health over a computer system that is powered by car batteries a portion of the time.

That carefully architected system has very little feedback and scant ability for evaluation. It is technically quite fragile, allows very few questions, has lots of opportunity for transcription errors, and has at least two languages involved (first report in Khmer, final report in English). We note carefully that much time and energy is consumed, everyone involved appears to be competent, dedicated, and wanting to do well, and yet the accuracy of even the best reports is suspect. Unfortunately, this is an current depiction of surveillance methods in one of the most likely outbreak areas on the planet.

Let us look briefly at the early stages of that submission sequence.

Data collection

This term encompasses active syndromic surveillance through direct collection as described above, plus confirmed case reports from the formal medical system, open and closed media surveillance systems, and environmental sensors.

There are many open media evaluation systems, including GPHIN, GHNN, Veratect, HealthMap, GEIS, and others. Each has limitations on sources, languages, identity extractions, software agents vs. human reviewers, filter bias, reporting cycles, and formats for submission to information consumers. Very few incorporate remote sensing data, medication purchasing sites (the first hint of the 1993 Milwaukee *Cryptosporidium* outbreak that sickened 400,000 people), or GIS layers, though all use geocoding for localization. None contain all of the information streams that might reveal a new disease outbreak, and there are strong statistical arguments that perhaps they shouldn't. As mentioned above, signal-to-noise ratios are an important part of assessment, and many existing data collection systems have optimized their retrieval to where they get reliable and reproducible information (a high SNR) for their consumers with the sources they are using and, for statistical reasons, should not add more. To add more sources without adequate evaluation would introduce noise and uncertainty in an epidemiological evaluation. That uncertainty could lead to unnecessary investigative deployments, unwanted public visibility, and socio-economic damage in the population under evaluation.

Related to the open-media analysis is remote sensing; an environmental assessment that collects physical parameters for processing with other, more conventional, media streams in algorithms that incorporate Bayesian inference and Poisson distributions for a determination of event likelihood according to the aptly-named "law of rare events."

Data transmission

In rural and remote areas, information flow starts with village elders and moves stepwise until reaching software developed for the World Health Organization in Geneva. At each stage the communication nodes can introduce inaccuracy or information loss. Methods for verification are few, and collection methods vary from memory jottings to paper ledgers, from cellular telephony conversations to interactive voice response (IVR), and from SMS messages to dedicated software over satellite links.

Data transmission is often viewed by developed countries as dependent upon electron flow, but that's not accurate in the developing world. The most effective intervention there for the improvement of deep-field baseline reporting can be as simple as a Ziploc bag protecting the paper ledger from the monsoon season, or a bicycle with weatherproof panniers. Through such seemingly small efforts records can remain readable during transport and transcription, and both reliability and accuracy of reporting may improve significantly.

Noise reduction

Noise reduction is the active improving of SNR, the signal-to-noise ratio, filtering out the extraneous, irrelevant, or unlikely data to focus on what's meaningful. It implies an ability to assess both relevance and credibility. Fortunately the mathematics of noise reduction in media analysis is quite well-developed, and capable analytical systems have evolved PhD programs (Georgetown, for example) around the furthering of the science. Two common methods include 1) the active removing (often by autonomous software agents) of source material unlikely to yield useful results prior to analysis, and 2) allowing raw data to be evaluated by pooled opinions of expert analysts (including, again, software bots) for likely credibility and value. Both retain the challenges of validation and verification, and the process of noise reduction itself risks the loss of genuinely useful data within the filtering.

Research

In medicine the ethics of patient research forces us to be mindful of those who have studied before us and that is an important concern with existing epidemiology and disaster research. Unfortunately, though, in studying outbreaks and disasters we often cannot see what happened. The primary materials are only rarely available for study.

We would advocate that every disease outbreak and every incidence of disaster have the primary materials associated with every phase of the investigation and response be kept in an open data set, curated by the originators, but open to others for study. The lack of availability of such primary data for re-analysis after the initial event is a concern and constraint that may be causing us to repeat avoidable errors in each response. Fortunately, restrictions on the access to original data are easily

altered by a little political and institutional will and we would advocate this Bellagio gathering include primary data accessibility standards in any position papers we produce.

Rather than elaborate further on the later stages of education, insight, decision support, and collective action, let us propose the following as goals.

In five years, by 2013, the informatics associated with disease surveillance could:

- be effectively blanketing the globe
- use visible and multi-spectral imaging of air, water and ground,
- use well-defined complex adaptive system assessments for emerging and resurgent disease risks,
- be informed by tested metrics, and
- be informed by careful research about lessons from historical events

Simultaneously, on the ground, practitioners could be taking full advantage of:

- robust cellular technology,
- far-forward pre-processing of data for district and provincial use,
- localized electronic libraries for reference and intervention,
- autonomous agents for decision support,
- collaborative software for response swarms across a mesh of information sources,
- appropriate incentives for government collaboration, community participation and support,
- iterative heuristics for constant improvements in diagnostic performance,
- integration of best practices from cultural anthropology and sociology to better understand the human factors that determine effective collective action for the populations we serve.

The result, we hypothesize, would be better early detection and better early response for diseases and disasters than we have in 2008. The technology is accessible, the policies are within reach, the educational requirements are well defined, the political will may soon be favorable, and the funding is not excessive. Such a system, in a typically invisible public health way, might both save lives and inspire second order efforts in local community resilience and nationally improved governance.

We challenge the attendees here at Bellagio to consider means and methods for pursuing such a goal.

Appendix 1:

Below is a description of the 31 data elements seen in the image above. These are the technical components we've identified as necessary for optimized biosurveillance.

As noted above, these five information domains:

1. Information flow
2. Mesh synchronization
3. Analysis
4. Decision support
5. Collaboration

contain the elements described below:

1. **Information Flow** requires:
    a. Forms design
        i. Forms design should allow data consolidation for optimized data collection by, and sharing between, organizations with responsibility for surveillance and response to priority infectious diseases.
    b. Shared ontologies
        i. The development of a common nosology of professional terms, including case definitions, is the basis for allowing disparate systems to communicate effectively with each other.
    c. Schema evolution
        i. Schema evolution (where schema is used in the sense of a plan or model for a database) is a design process that ensures that modifications to the structure of the database do not result in data loss. This implies flexibility in the database architecture, allowing users to modify a database over the years to fit new requirements, and it also mandates backward compatibility.
    d. Translation
        i. As discussed above, linguistic, ontological, and cultural.
    e. Deep-field collection
        i. There has to be a reliable and reproducible collection system that allows communication in both directions, including feedback to the field. There are several excellent methods for collection available, including EpiSurveyor, GeoChat, Voxiva, and Los Alamos Avian Tracker, and more are appearing annually. Reliable feedback methods from the Ministry back to the affected population are more difficult to identify in our surveys, as are cross-sector or intra-sector communications within the same geographic locale.
    f. Geocoding

i. In the modern world, epidemiological information should contain reliable geographic coordinates and be plotted on a map, preferably with any other GIS information layer of possible relevance (e.g. temperature, humidity, previous events near that location, rainfall variance, transportation hubs, phylogenetic variation, etc).

2. **Mesh synchronization** requires:

    a. Storage abstraction

        i. This is a term to describe the persistence of information unaltered across systems, regardless of the application using the information.

    b. Offline work

        i. Most of the data collection done in public health is done with no access to the Internet, and often not even to cellular telephony. Systems intended for data acquisition should be designed to accommodate disconnected users at the edge of the infrastructure, yet synchronize seamlessly with Internet systems when connectivity is re-established. Such techniques are commonplace within industry (Allianz-Indonesia, the US Federal Aviation Administration, Steelcase Corporation, Monsanto Chemicals, Pennsylvania Office of the Attorney General, the European Railway Agency, Statoil-Norway, and many more) but rarely integrated into public health tools.

    c. Conflict resolution

        i. This is a term for synchronization techniques, ensuring that two people collecting medical data offline through the day update each other gracefully when they re-establish connectivity. The mathematics is difficult, but understood, and software techniques for conflict resolution have been maturing for more than a decade.

    d. SMS integration

        i. As above, cellular telephones, using SMS text messaging to minimize bandwidth and cost, are the most ubiquitous and resilient form of electronic communication available. We consider incorporating SMS to be a design mandate in any primary reporting tool.

    e. Security

        i. We are conveying health information, so those international standards apply, and we are also conveying information that might, if made public, result in social and economic disturbance, regardless of the veracity of the report, so encryption is required.

    f. Identity

        i. We consider it imperative for quality control that any submission in any official epidemiological reporting system be identified by source, date, and time for accountability. Changes may be allowed, but the changer must be identified just as clearly as the original creator.

    g. Adapters and transformers

i. Our experience shows that when peer colleagues join us in the field during a response, they bring all manner of hardware and software with them. Although often designed for information management, many of the tools may not work well for sharing information with tools brought by other organizations. The reasons for organizational choices in communications tools may be educational, financial, nationalistic, or any number of other causes and they are often not much affected by externalities. In our experience IT resources are most commonly an internal decision, little affected by outside considerations. Mandating a particular platform or application across agencies and organizations (even if some superagency had the authority) is therefore somewhat arrogant, impractical, and unlikely to succeed. Our opinion, though, is that collaboration should be strongly encouraged at every opportunity, with information shared to the greatest degree acceptable. We should, therefore, work to find or develop tools to permit information sharing across all platform and application boundaries (from spreadsheets to databases to GIS layers to document handlers to communications devices to presentation software to statistical programs), with negligible additional technical support from the users except for establishing the levels of permission.

3. **Analysis** requires:

    a. Data fusion

        i. Data fusion is the technical framework for the alliance of data originating from different sources for a common purpose. An example might be data from temperature and humidity sensors, satellite multispectral images, rainfall totals, and freshwater GIS maps, combined to assess mosquito breeding areas.

    b. Anomaly detection

        i. This is the ability to identify, at the earliest moment, a sign that signifies something different from baseline. This requirement drives the development of fine-grained data collection systems and the defining of algorithms for separating signals from noise.

    c. Complex Adaptive Systems modeling

        i. This refers to the need to collect all of the likely streams of information that could alter our measure of public health risks, then assess them for patterns in their complexity, looking for evolutions in the system that might pose a threat. The streams required for analysis include economies, ecologies, cultures, weather, social organization, molecular epidemiology, and so forth.

    d. Network analysis

        i. This is the mapping and measuring of flows between people, groups, organizations, animals, computers, and other information processing entities. This is, naturally, a necessary step in optimizing predictive value, completeness, security, and timeliness in outbreak response.

    e. Text mining

        i. Information extraction to discover non-obvious facts and relationships, usually used in public health to describe open-media analyses for issues of biological significance. Well-known organizations performing text mining in public health include Veratect, GPHIN, GHNN, Healthmap, GEIS, Google and others.

    f. Spatio-temporal analysis

i. The evaluation of events across locations and time, giving historical context and the possibility of detecting patterns and causes, perhaps leading to the prediction of future events.

g. Sensor integration

i. As in data fusion above, the process of accepting multiple sensor feeds into a data stream, incorporating their value into the information mesh, often displayed as a GIS layer.

4. **Decision support** requires:

a. Geospatial visualization

i. The process of viewing geospatially tagged information on maps or satellite imagery. It includes photogrammetry, digital terrain modeling, Photosynth graphic capabilities and more. It's used for disease vector mapping, cluster sampling, and intuitive grasping of information that was originally in tabular or graphical form and so difficult to assess.

b. Autonomous agents ("bots")

i. In the software sense, these are pieces of code that act for a user or another computer program in a relationship to data. They act "on behalf of" and so have the authority to make decisions on whether an action is appropriate. As agents, they activate themselves, and as "autonomous" agents, they are capable of independently modifying the way in which they achieve their objectives. Based on rules, these agents have persistence, autonomy, social ability (e.g. they can collaborate on a task with other systems) and reactivity, in that they can perceive the context in which they operate and react to it appropriately.

ii. In the path from event recognition, to collaborative sensemaking, to collective action, bots can provide users within collaborative groups an awareness of activities elsewhere. Such awareness of the global state of the system may inform decisions at the local level, a phenomenon recently defined as "immergence". Despite the futuristic impression this explanation might give, autonomous agents are common and extremely useful. In open media surveillance and in collaboration software, they lighten the human workload efficiently and accurately.

c. Predictive modeling

i. The process by which a model is created or chosen to try to predict the probability of an outcome. There are a number of mathematical techniques used for the development of the model and they are well-characterized and commonly available. InSTEDD, for example, uses Support Vector Machines, one of the techniques, in our RNA toolset.

d. Distributed workflow

i. A process for optimization of a task whose parts are separated in time and space. An example is the creation of software between teams in India, Amsterdam, and Argentina who, between them, use tools to parse the problem, distribute sections, track versions, and manage quality control, all while separated by time zones and geography. It might also refer to outbreak notifications that require specialists in several regions and time zones to contribute to the evaluation in a manner that allows normal workdays and the accumulation of opinion in a central location until the evaluation is adequate to allow a decision.

e. Alerting

i. As expected, the provisioning of safety warnings when an important item is reported. Alerts should be both human and agent-based and should provide warnings over any mode of communication used by those who must be informed: SMS messages, pre-recorded telephone messages in native languages, loudspeakers, flashing lights, emails coded at highest priority, and more.

f. Report generation

i. Report generation is a necessary aspect of surveillance, but the process can now be automated to a large degree. Reports can (and should) also be multi-modal, with images, maps, video, 3-d modeling, audio, charts, and other media used whenever they would enhance understanding.

5. Collaboration

a. Directory federation

i. This describes the ideal assembly of Who-What-Where lists, allowing professionals across and within specialties to find each other. The intent is to have cross-subscription between each of the many instances of these lists (which mushroom in emergencies) so that anyone on one list can find anyone on another. That requires both technology and permission, and the permission probably requires incentive. A discussion of all three – technology, permissions, and incentives, is underway.

b. Social networking

i. A method for connecting individuals through other individuals, now often through a social networking site like Facebook, Linked In, Ning, MySpace, Orcut and others. There is a trust model that operates within social networking, implying, in the most general sense, that a friend of my friend is my friend. There appears to be genuine value in social networking and, in an edition of New Scientist in March 2008, Facebook was described as "very useful" during the Virginia Institute of Technology campus shootings and during the San Diego fires in 2007.

ii. We need a collaborative framework in which autonomous agents facilitate introductions between users, and the sharing of sensitive, yet critical, information between groups, without breaking the circle of trust. We think that the social graph underlying such social networking services may be used as a filter for information sharing in the transition from isolated groups collaborating to more global collective action and InSTEDD is testing that hypothesis.

c. Virtual teaming

i. A description of teams that operate across time, distance, and organizational boundaries, occasionally working for years together and never meeting in person. The tasks are often ephemeral but the teams may be permanently established.

d. Reliable messaging

i. There are several senses intended in this term, starting with the risk messaging back to a vulnerable population, including both method and content. It also incorporates the sense of a virtual team swarming around a problem for analysis and ensuring that their intra-team communication is effective.

e. Social metadata

i. This is the value added by participants in a collaborative evaluation. From the original item, perhaps a measles outbreak in northern Vietnam, the social metadata could include a measles case definition, the weather recently, the transportation and social hubs common to this patient and his culture, any

historical cases of measles in this area, the number of children in the immediate area, and the date and results of the last measles vaccination campaign. All of that is social metadata – information from others that augments the original item and improves the contextual understanding of the item for the decision makers.

END