OPINION

Specimen collections should have a much bigger role in infectious disease research and response

Diane DiEuliis[a], Kirk R. Johnson[b], Stephen S. Morse[c], and David E. Schindel[b,1]

When public health officials become aware of the first signs of a disease outbreak, they need to determine a few critical things as quickly as possible. What's the disease agent? How did it get here? How does it spread and how can it be contained? Has it been seen before? If so, what was the approach and how well did it work?

The answers to these questions are often not only elusive but complex. We believe that an array of helpful investigative tools have been overlooked. The answers to these and other critical questions may be available "off the shelf": the shelves of scientific collections in research institutes and museums. Taking better advantage of such collections could offer public health agencies at all levels—local, state, national, and international—a much larger toolkit for characterizing, mitigating, and predicting emerging infectious diseases. In doing so, scientific collections could find new users, new impact, and new sources of support. Public health first-responders and scientific collections might not seem like obvious allies, but they can benefit from deeper collaboration. We suggest that researchers, public health officials, and others should consider collections related to infectious diseases as ready resources that should be used routinely and supported collaboratively. Government agencies and private donors should create new interdisciplinary mechanisms for better funding of collections that serve these new uses, as well as support for long-term curation and preservation of these important legacy collections.

Field veterinarians collect a bat blood sample for pathogen surveillance (*Upper Left* and *Upper Right*) in conjunction with the US Agency for International Development's Emerging Pandemic Threats PREDICT project, a collaboration between the University of California, Davis and Tanzania's Sokoine University of Agriculture. Specimens preserved in alcohol represent ~97,000 of the 590,000 specimens in the Division of Mammals at the National Museum of Natural History in Washington, DC. (*Lower Left*) Frozen tissue and DNA samples are preserved with liquid nitrogen (*Lower Right*) at the Museum Support Center in Suitland, MD. Images courtesy of (*Upper Left* and *Upper Right*) the University of California, Davis, One Health Institute, Davis, CA, and (*Lower Right*) Donald E. Hurlbert (Smithsonian Institution).

Collecting Clues

By "scientific collections," we mean the physical objects obtained for research in an extremely broad range of disciplines, from archaeology and anthropology to biology, biomedicine, earth sciences, and applied fields, such as agriculture and technology. Some portion of the samples and specimens obtained, managed, and used by researchers may, at some point, be turned over to collection managers for long-term preservation in what have been termed "institutional collections" (also sometimes referred to as "archival" or "reference" collections). Once accessioned into an institutional collection, the samples and specimens are made available to qualified researchers for study.

Institutions like the Smithsonian come to mind when collections are mentioned, but there are tens of thousands of institutions with collections (museums, hospitals, universities, veterinary and medical colleges, government agencies, zoos, botanical gardens, even private companies, such as the American Type Culture Collection in Manassas, VA). Collectively these institutions contain tens of thousands of different collections (e.g., preserved plant, animal,

[a]Policy and Planning, Office of the Assistant Secretary for Preparedness and Response, US Department of Health and Human Services, Washington, DC 20201; [b]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560; and [c]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032

Author contributions: D.D., K.R.J., S.S.M., and D.E.S. wrote the paper.

Any opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the National Academy of Sciences.

[1]To whom correspondence should be addressed. Email: schindeld@si.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1522680112

and microbial specimens stored according to their taxonomy, living organisms, frozen tissues and DNA, living cell lines). Unfortunately, collections too often enter the public's consciousness only when they're in financial distress and facing possible closure (1), or when institutions have mismanaged important samples (2, 3), or violated biosafety protocols (4–6). Too seldom does the public learn of cases where these collections provide the critical evidence for solving a problem facing science and society. A recent workshop organized by Scientific Collections International (7) and held at the Smithsonian Institution examined one such problem: combatting emerging infectious diseases (EIDs).

Growing Threats

The past decade has seen the outbreak of several destructive infectious diseases (e.g., SARS, H1N1), the most recent being the devastating Ebola outbreak in West Africa (8, 9). The public's consciousness of these outbreaks, guided by the media, has focused on disease transmission through interpersonal contact. Travelers may be asked if they have visited certain countries and news reports show infrared screening at airports, looking for passengers with fevers. News coverage of the Ebola crisis in the United States has concentrated on transmission of the disease from patients to doctors, nurses, and other caregivers who might then carry the infection internationally by plane travel.

However, public health agencies and disease researchers know that disease transmission between wild or domesticated species and human populations may be a much greater risk than international travel. One study reported that about 60% of the 1,400 known pathogenic species are zoonotic (i.e., they can be passed between humans and other animal species), and 75% of the 175 species that are responsible for serious disease outbreaks are zoonotic (10). Another 17% of human diseases were introduced by insects or other types of vectors (11).

EIDs are therefore more properly considered to be characteristics of complex ecosystems that involve human society, livestock herds, animals and their food products in the marketplace, and populations of wild animals, all set in the context of changing climate, weather, and land-use patterns. Scientific collections are becoming critical sources of information on how diseases pass among the components of these ecosystems. These collections are widely distributed across diverse organizations, including natural history museums, research centers devoted to pathogens, reference collections of parasites and livestock, veterinary and medical colleges, hospitals, public health agencies like the CDC, Food and Drug Administration, and NIH, and military organizations, such as the US Army Medical Institute of Infectious Diseases in Frederick, MD and the Joint Pathology Center in Silver Spring, MD. Their contents are not widely known among researchers, especially across disciplines.

Consider, for example, the 1918 Spanish influenza outbreak that claimed ~50 million lives, and the periodic emergence of similar flu pandemics since then (12). The origins of the disease and the relationship between the 1918 pandemic and subsequent outbreaks have been studied by comparing viral RNA sequences from modern patients, livestock (e.g., swine and fowl), and wild birds with samples preserved in the Joint Pathology Center and the Smithsonian's National Museum of Natural History. This enabled researchers to reconstruct the 1918 influenza virus and its subsequent history (13, 14). Reference collections such as these are our only window into the past geographic distributions of different disease agents and the changes they have undergone.

In the early 1990s, natural history museums in the southwest United States were drafted in the effort to contain hantavirus outbreaks (15). Small rodents were suspected to be the agents by which the virus was spread. Collections like those in the Museum of Southwest Biology at the University of New Mexico had extensive rodent collections for ecological and evolutionary studies. Museum of Southwest Biology researchers had been doing holistic sampling that preserved museum specimens for traditional taxonomic study (stuffed hides, skeletal, teeth), as well as forward-looking samples that would support other fields of study (e.g., endo- and ecto-parasites, gut contents, and frozen tissue samples for DNA and RNA analyses). The frozen tissue samples revealed when and where the virus appeared and how it spread among rodent populations. Analyses of the museum collections allowed public health officials to predict and anticipate new outbreaks among human populations.

A less-successful outcome offers a cautionary tale. Lassa fever is relatively common in the west African countries of Sierra Leone, Liberia, and Guinea, and patients with fever symptoms are often given this initial diagnosis. However, Lassa is a misdiagnosis in perhaps 60–70% of these cases and the Lassa-negative cases turn out to be a variety of other EIDs. Not every Lassa-negative infection is identified, leaving many of the region's EIDs uncharacterized. One study that retrospectively tested samples from the undiagnosed cases between October 2006 and October 2008 found Ebola-positive sera from this period, a full six years before the major West African outbreak (16). No subsequent collections were made of human blood samples or the bats and other rodents suspected of carrying Ebola, even after a 2012 report that Ebola antibodies were found in Ghanaian bats sampled a year before the 2008 cases (17). This failure to create reference collections of samples from patients and possible wildlife reservoirs left no resources for

Collections can provide short-cuts to public health responders looking for the origins and distribution of disease agents, but only if the collections are accessible and well documented.

predicting and mitigating the next emergence of Ebola in the region.

The cost of conducting several years of field studies leading to the collection and analysis of small mammals is typically several hundreds of thousands of dollars. Maintaining these collections costs a small fraction of that amount. In contrast, the World Bank mobilized more than $1.6 billion for Ebola recovery and estimated that the region's gross domestic product would lose $2.2–7.4 billion over the short-term (18). The story of the 2014 Ebola outbreak might have been very different if relatively small investments had been made to collect field samples after the initial discovery and to identify the local distribution of Ebola viruses from those collections. Such data would also have helped to put speculation about new mutations in context.

Measures to protect human populations from Guinea Worm have been extraordinarily effective once all its different growth stages and host species were sampled, preserved, documented, and analyzed. The pathways of this and other tropical diseases, such as dengue fever, elephantitis, and schistosomiasis (19) through different hosts can only be understood by making and studying populations of potential host and vector species, both wild and domesticated. Access to collections and the results of their analyses will be the rate-limiting factor in understanding and interrupting cycles of multispecies disease transmission.

Applying the Lessons Learned

Several lessons emerge from these case studies. First, collections can provide short-cuts to public health responders looking for the origins and distribution of disease agents, but only if the collections are accessible and well documented. Second, undiagnosed diseases often seem to disappear with the last reported case but they may be thriving in a nonhuman reservoir species, waiting to re-emerge. The cost of making, documenting, and maintaining collections following a short outbreak is miniscule relative to the cost of responding to a re-emergence without the benefit of information that collections can provide. Third, and perhaps most critical, these benefits can only be realized through a new culture of collaboration between the public health sector and the relevant research communities that rely on scientific collections.

We propose a new and interdisciplinary enterprise that will produce new collections of organisms, microbes, tissue and fluid samples, DNA extracts, and will leverage existing collections to the greatest degree possible. The digital data and metadata associated with these collections (e.g., images, traits, DNA sequences, and public health information) will need to be standardized and interoperable across fields and countries so that information is available anywhere, anytime. What sets this enterprise apart from other collecting initiatives is collaborative planning, funding, and implementation between infectious disease researchers and colleagues and institutions in other fields. New efforts are needed to bring together specialists in the public health sector and the research domains that will collect, manage, and analyze collections relevant to emerging infectious diseases (e.g., wildlife biologists, parasitologists, livestock and veterinary scientists, microbiologists). Funding agencies and international programs like the One Health Initiative have critical roles to play by providing interdisciplinary support without the barriers that traditionally separate these fields.

Public health agencies have the daunting task of preparing for the unknown, but we believe that strategic investments in scientific collections help to inform their decisions. Acting by themselves, natural history museums, other collection repositories, and public health agencies are not equipped to meet these challenges. Prioritizing, designing, and executing collecting programs will require new partnerships and interdisciplinary funding programs. The same will be true for documenting, preserving, and analyzing samples, and for making this information available.

In calling for increased and interdisciplinary funding for collections with importance for public health, we are mindful of all of the life-cycle costs associated with long-term sample preservation. Public health researchers, research curators, and collection professionals face many decision-points that involve cost-benefit trade-offs. Consider three different types of collecting programs. Relatively inexpensive "just-a-look" monitoring programs can collect and analyze small numbers of samples to look for evidence of disease agents in regions suspected of being at risk. "Just-in-case" projects can collect specimens and preserve tissue and blood samples but delay genetic or other testing, thereby reducing costs. "Just-in-time" projects can be launched on short notice following an outbreak. They involve immediate testing of samples and are the most expensive of these three program designs.

In all cases, funders and researchers must make tactical decisions. Should whole organisms be collected and preserved or would digital pictures of the organism associated with tissue and fluid samples be sufficient? Is a whole voucher specimen needed for every tissue sample or would a few representatives be sufficient? Do tissue samples need to be retained once DNA has been extracted? Once sequences have been archived in GenBank, is there value in preserving DNA extracts? At each point, the potential for future use (both for public health applications and for other biological research) must be weighed against the costs of long-term preservation. New funding for collecting and collections should carry with it the expectation that informed decisions will be made about what to collect, preserve, and retain.

Acknowledgments

The authors acknowledge the participants in the Scientific Collections International workshop (7) for many of the ideas and insights discussed here.

1 Kemp C (2015) Museums: The endangered dead. *Nature* 518(7539):292–294.
2 House of Representatives (2008) Hearing Charter: Biobanking: How the Lack of a Coherent Policy Allowed the Veterans Administration to Destroy an Irreplaceable Collection of Legionella Samples. Available at https://science.house.gov/sites/republicans.science.house.gov/files/documents/hearings/090908_charter.pdf. Accessed November 25, 2015.
3 Ledford L (2008) Investigation into lost bacteria collection raises concerns about biobanks. *Nature Online*. Available at www.nature.com/news/2008/080910/full/news.2008.1096.html. Accessed November 25, 2015.
4 Reardon S (2014) NIH finds forgotten smallpox store. *Nature Online*, 10.1038/nature.2014.15526.
5 Reardon S (2015) US military accidentally ships live anthrax to labs. *Nature Online*, 10.1038/nature.2015.17653.
6 Government Accountability Office (2014) High-containment Laboratories: Recent Incidents of Biosafety Lapses. GAO-14-785T. Available at gao.gov/assets/670/664799.pdf. Accessed November 25, 2015.
7 Scientific Collections International (2015) Scientific Collections and Emerging Infectious Diseases. Available at scicoll.org/scicollpubs/EID_2015March.pdf. Accessed November 25, 2015.
8 Morse SS (1995) Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1(1):7–15.
9 Morse SS, et al. (2012) Prediction and prevention of the next pandemic zoonosis. *Lancet* 380(9857):1956–1965.
10 Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 356(1411):983–989.
11 WHO (2014) Vector-borne diseases. Factsheet no. 387. Available at www.who.int/mediacentre/factsheets/fs387/en/. Accessed November 25, 2015.
12 Taubenberger JK, Morens DM (2006) 1918 Influenza: The mother of all pandemics. *Emerg Infect Dis* 12(1):15–22.
13 Tumpey TM, et al. (2005) Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* 310(5745):77–80.
14 Taubenberger JK, et al. (2012) Reconstruction of the 1918 influenza virus: Unexpected rewards from the past. *MBio* 3(5):e00201–e00212.
15 Yates TL, et al. (2002) The ecology and evolutionary history of an emergent disease: Hantavirus pulmonary syndrome. *Bioscience* 52(11):989–998.
16 Schoepp RJ, Rossi CA, Khan SH, Goba A, Fair JN (2014) Undiagnosed acute viral febrile illnesses, Sierra Leone. *Emerg Infect Dis* 20(7):1176–1182.
17 Hayman DTS, et al. (2012) Ebola virus antibodies in fruit bats, Ghana, West Africa. *Emerg Infect Dis* 18(7):1207–1209.
18 World Bank (2014) The Economic Impact of the 2014 Ebola Epidemic: Sort- and Medium-Term Estimates for West Africa. International Bank for Reconstruction and Development. Available at https://openknowledge.worldbank.org/bitstream/handle/10986/20592/9781464804380.pdf. Accessed November 25, 2015.
19 Molyneux DH (2004) "Neglected" diseases but unrecognised successes—Challenges and opportunities for infectious disease control. *Lancet* 364(9431):380–383.