ESTABLISHING AN NIH-WIDE GEOSPATIAL INFRASTRUCTURE FOR MEDICAL RESEARCH: OPPORTUNITIES, CHALLENGES, AND NEXT STEPS

Report of the AAG-NIH Workshop on
Geospatial Infrastructure for Medical Research, 2011

INTRODUCTION

The need for spatial and spatiotemporal analysis arises in numerous areas of NIH biomedical and public health research, and NIH Institutes and Centers increasingly recognize the importance of geographic context in this research. The past two decades have seen dramatic increases in the use of geographic theories, data, methods, and tools to help respond to this need. For example, Geographic Information Systems (GIS) and spatial modeling are being used in research on the epidemiology of cancers (NCI); social epidemiology research related to drug abuse and treatment (NIDA), studies of gene-environment-health interactions (NIEHS); heart disease, stroke, asthma, and COPD (NHLBI); infectious-disease transmission, ecology, and spread (NIAID and Fogarty Center); understanding the relationship between UV radiation, vitamin D levels, and MS prevalence (NINDS); small-area analyses of pain and access to care for pain-related conditions (NINDS); and on themes related to global health and health disparities.

While such examples demonstrate great progress in recent years in developing GIS, geocoding services, mapping, and associated standards, challenges still abound. These include the lack of interoperability among proprietary systems, longitudinal variation in data collection, difficulties of sharing inadequately documented data, issues of confidentiality of location-specific data, lack of understanding of the basic concepts of spatial and spatiotemporal data and analysis, and redundancy of effort and investment. A comprehensive or uniform strategy to incorporate geographic context across the breadth of biomedical and public health research at NIH does not yet exist.

Furthermore, most health-science applications do not take full advantage of the latest developments in spatial and spatiotemporal data analysis and modeling, or the new types of geographic data and computing resources that are becoming available. These developments are related to: 1) the explosion of real-time, spatiotemporal data from GPS-enabled devices, distributed environmental sensor systems, satellite remote sensing, and (potentially) from geographically tagged electronic medical records; 2) development of new tools and methods for analyzing spatiotemporal data, including methods of geovisualization, dynamic spatiotemporal modeling, and modeling of human mobility at scales ranging from the everyday to the life course; and 3) advances in computing technologies, service-oriented architectures, and cyberinfrastructure that are fueling the growth of distributed and collaborative services known as the geospatial web.

Such challenges are common across the diverse Institutes and Centers of NIH, cutting across the social and biomedical sciences. However, NIH-funded research incorporating geographic approaches has thus far been carried out primarily on a project-by-project basis, mitigating the potential advancement of geographic data, methods, and theories across NIH divisions. Very substantial scale economies, not to mention opportunities for innovative and collaborative research discoveries, can be achieved by addressing them collectively. While many institutes have made substantial investments in spatial data and tools, a collective approach through a common infrastructure would offer significant advantages.

To evaluate the potential development of an **NIH-wide geography and geographic information infrastructure** ("geospatial infrastructure") to support basic biomedical research and public health applications, the Association of American Geographers (AAG), the National Cancer Institute (NCI), and the National Institute on Drug Abuse (NIDA) recently co-sponsored a highly-successful workshop. Participants included senior scientists from across the National Institutes of Health (NIH), leading researchers in GIScience, NIH-funded researchers who use geographic theory and methods in their research, and industry experts on geographic technologies. The workshop was held on February 22-23, 2011 at NIH facilities in Rockville, Maryland. This report presents the *key ideas* along with a series of proposed *next steps* that emerged from workshop presentations and discussions. In addition, Appendix A includes a summary of the workshop, Appendix B lists workshop participants, and Appendix C details the workshop agenda.

KEY IDEAS

An NIH-wide geospatial infrastructure should be broadly conceived to encompass technology, architecture, integrated and interoperable spatiotemporal databases, metadata and standards, analytical methods and tools, visualization, data access and privacy protocols, and training and capacity building in geographic theory and analysis. The following key ideas synthesize the comments and suggestions shared by workshop participants, and provide a broad framework in support of an NIH-wide geospatial infrastructure.

1. **Integrate spatial activities across institutes.** Integrate spatial data and analysis into processes for research, discovery, and health serves delivery across NIH. This type of an initiative – "Spatial @ NIH" – could be very useful toward bridging and integrating spatial activities across the institutes and facilitate inter-institute communication and pooling of resources. Such an initiative could demonstrate the value and usefulness of integrating geographic analysis across all institutes, support the missions and strategic plans of each institute, and provide a strategic and essential foundation to an NIH-wide geospatial infrastructure.

2. **Establish goals for an NIH-wide geospatial computing infrastructure.** There is a clear need for a common geospatial computing infrastructure for NIH and NIH-funded researchers. This infrastructure would increase efficient use of research funding since principal investigators would not have to replicate the parallel establishment of their own geospatial infrastructures. For example, such an infrastructure could include a

searchable spatial-data sharing platform that incorporates research on data confidentiality mechanisms. An NIH-wide geospatial infrastructure needs to be forward-looking and adaptable to rapid changes in GIScience research and technology (such as real-time data collection and analysis, social media, crowd-sourcing, electronic health records, individual sensors), cyber-infrastructure, cloud computing, and related technologies. Efforts to build an NIH-wide geospatial infrastructure should be informed by both the successes and failures of other existing large, scale institutional models.

3. **Incorporate spatial context in health data and research.** Health behavior and outcomes are likely influenced by social and physical environmental contexts that operate at different geographic scales. Clear methodological frameworks and methods for capturing and quantifying the effects of these multiple contexts at numerous scales need to be established. A research-oriented geospatial infrastructure that incorporates *multi-scale* context will create opportunities for new research hypotheses and discoveries linking environment, behavior, and health outcomes through *space and time* in ways that have never been possible before. As a component of successful infrastructure development, established (and future) national health surveys need better localized geo-sampling schemes matching research locations with disease. Sampling survey design needs to consider representativeness in terms of the contextual characteristics of places as well as individual characteristics.

4. **Develop education initiatives for spatial thinking and geographic methods.** An NIH-wide geospatial infrastructure should include active education efforts, for conceptual spatial thinking as well as for the specific methods to enable application of geographical analysis in health research. Efforts should be made to document and widely disseminate information about existing tools, methods, and best practices. It is important that education efforts focus on more than simply the capabilities of GIS software programs but, rather, on the full range of concepts and methods of geography and spatial thinking, and utilize appropriate training opportunities already in place at NIH, such as the R25 Research Education grant mechanism.

NEXT STEPS

This report, which incorporates review comments from the research community, will be presented to the leadership across NIH to begin planning specific actions in and across institutes. The following proposed next steps have been categorized as short term or long term to capture general ideas about their scope and feasibility.

Short Term

- **Submit articles to *Science* and other peer-reviewed biomedical and geographic journals.** Write and submit articles for *Science* and other relevant peer-reviewed journals describing the need for, value of, and progress toward developing an NIH-wide geospatial infrastructure. Such articles could clearly articulate how a common geospatial infrastructure will facilitate the advancement of important scientific questions in biomedical research, and should include examples of research that illustrate the impact of spatial thinking on health outcomes.

- **Inventory spatial research activities across institutes**. Conduct a review of NIH grants to identify all that use geospatial tools or whose study goals include geographic aspects of health. Develop a map of spatial activities across institutes. This could be modeled after a similar effort at NSF that examined the use of spatial language in NSF award abstracts. Identify research gaps as well as existing geospatial infrastructure that could serve as foundation for an NIH-wide infrastructure.

- **Document and disseminate information about existing tools, methods, and best practices.** Compile an NIH-wide library of existing GIS tools, software extensions, and best practices; make them available online or link to existing online sources; and broadly disseminate information about their availability. Facilitate the development and use of various tools, including ArcGIS extensions for health. This effort could be modeled after the process the EPA created to distribute environmental exposure extensions. Develop and provide documentation about best practices for applying geographic methods to health research. This should include methods, guidelines, and standards for protecting the confidentiality of geocoded data on individuals.

- **Establish requirements for an NIH-wide geospatial computing infrastructure**. Work with potential users to develop a set of functional requirements for an NIH-wide geospatial computing infrastructure. Work with other government agencies that are or have been involved in building geospatial infrastructures. Understand what seems to work well and what does not. Identify opportunities to leverage work being done by others. Engage the research community across all of the institutes in the definition of an NIH-wide geospatial infrastructure.

- **Support geographic data in electronic health records**. Encourage the addition of location variables/geocodes to emerging standards for an electronic health records system, creating a foundation to conduct spatiotemporal research. These standards should include, for example, fields for multiple addresses to allow the collection of residential histories.

- **Expand geospatial education.** Identify opportunities to incorporate geography and GIScience sessions (awareness) into NIH meetings across a wide range of institutes, continue to expand health research sessions at geography and GIScience meetings, and encourage geographers and GIScientists to serve on NIH review panels. Develop an NIH training track (similar to the NSF track) to provide technical assistance to geographers to support the development of training awards (NRSA, K awards, dissertation).

- **Improve spatial aspects of PubMed**. Work with the National Library of Medicine to include more geographic journals with articles on health-related geospatial methods, maps and data; add spatial MeSH terms and headings; and include study area geography in search capabilities.

Long Term

- **Build consensus and develop a strategic plan.** Engage an outside expert for a period of time (6-12 months) to drive initial efforts towards an NIH-wide geospatial infrastructure, such as a senior, highly respected geographer who is capable of developing a consensus statement of needs across NIH users and others; and also facilitate implementation of the ideas. In addition to integrating spatial activities across institutes, an outcome of this effort should be a strategic plan that would include funding requirements and alternatives. The plan should establish needs and opportunities for research funding or additional workshops in areas such as (but not limited to): distributed computing and geo-computational needs for large spatiotemporal data sets; identifying specific research needs related to spatiotemporal analysis and health research; analysis of longitudinal data; access to data and privacy concerns; ontology and common language; and trends in research and technology (e.g., cyber-infrastructure, geospatial web, social media, electronic health records). This activity could grow into a Geography Division for NIH, an Office of the Geographer, a Geographic Information Officer, or a Center for Spatial Analysis at NIH.

- **Establish a community of users**. Establish and support a community of users drawn from geography/GIScience and health/biomedical research. The community would support mentoring of early career researchers as well as provide a vehicle for NIH-funded researchers to communicate the importance of a common geospatial infrastructure to NIH leadership.

- **Establish an NIH-wide geospatial infrastructure**. Collaborate across institutes, agencies, and organizations to develop the geospatial infrastructure (leverage technology advances), reduce research burden, increase cost sharing, combine data, build capacity, and deliver services and interventions. Draw support and expertise from the geography/GIScience and health/biomedical communities and build on current efforts.

CONCLUSION

This initiative offers the potential to open new doors for geographic research and discovery, in collaboration with both NIH intramural and extramural biomedical scientists and with related public health researchers. For geographers and biomedical researchers alike, it also holds real promise for making a meaningful difference in the health and lives of people around the world.

REPORT PREPARED BY:

Association of American Geographers

Douglas Richardson
Jean McKendry
Michael Goodchild
Mei-Po Kwan
Sara McLafferty

National Cancer Institute

Zaria Tatalovich
David Stinchcomb (Westat)

National Institute on Drug Abuse

Bethany Deeds

Appendix A
Workshop Summary

Presentations at the workshop included an overview of current geospatial activities at NIH institutes, perspectives from the GIScience research community, views of extramural NIH researchers on the potential of a common geospatial infrastructure to advance science, and system architectural considerations for a geospatial infrastructure. Breakout groups during the workshop focused on identifying common needs, key challenges, and implementation alternatives.

Participants in the workshop agreed that developing a broader and deeper geospatial infrastructure throughout NIH for biomedical research is needed. An NIH-wide geospatial infrastructure should be broadly conceived to encompass technology, architecture, integrated and interoperable spatiotemporal databases, metadata and standards, analytical methods and tools, visualization, data access and privacy protocols, and training and capacity building in geographic theory and analysis.

Workshop participants felt it was important to clearly establish the scientific value of geospatial investments. The discussions highlighted numerous benefits of geography and GIScience to NIH's health research programs. Examples of the benefits of a large-scale geospatial infrastructure to health and biomedical researchers include: generation of research hypotheses through discovering spatial patterns and relationships; ability to make causal inferences by including the temporal dimension in spatial analysis; increased ability to understand gene-environment interactions and their role in disease occurrence; ability to advance mobile health systems by incorporating real-time GPS/GIS technologies; and the potential to integrate and link individual level information in other major health data sources with the context and neighborhood of those individuals.

During the workshop, participants also discussed challenges to the implementation of such an ambitious project. These challenges include dealing with locational privacy and confidentiality; developing and disseminating geospatial tools specific to the needs of health and biomedical researchers; and providing training and education in spatial thinking and the use of geospatial methods to health and biomedical researchers. With the growing availability of rich multi-level data sources on individuals and their neighborhood contexts as they move across space and time, issues of scale, large datasets, and computational capacity need to be considered. Developing a robust distributed computing architecture (including cloud computing) for an NIH-wide geospatial infrastructure that meets the needs of all interested parties and takes advantage of other national efforts will also be challenging. Finally, the development a common, shared language and a set of ontologies for geospatial methods and data that is shared by the diverse disciplines involved in biomedical research will be critical to foster interdisciplinary collaboration.

Participants recognized the importance of having a forward-looking strategy in developing an NIH-wide geospatial infrastructure, being mindful of new and emerging technologies including, for example, the geospatial web, social media, new information from electronic health records, real-time health monitoring, and developments in sensor and location-aware technologies.

Appendix B
Workshop Attendees

David Balshaw (NIEHS)

David Berrigan (NCI)

Regina Bures (NICHD)

Jarvis Chen (Harvard University)

Wilson Compton (NIDA)

Paul Courtney (NCI)

Ellen Cromley (University of Connecticut School of Medicine)

Bob Croyle (NCI)

Bill Davenhall (Esri)

Bethany Deeds (NIDA)

Brenda Edwards (NCI)

Mike Goodchild (University of California, Santa Barbara)

Paul Gruenenwald (Pacific Institute for Research and Evaluation)

Colette Hochstein (NLM)

Geoff Jacquez (BioMedware – Ann Arbor, MI)

Mei-Po Kwan (Ohio State University)

Gene Lengerich (Penn State)

Amy Lobben (University of Oregon)

Jonathan Mayer (University of Washington)

Jean McKendry (AAG Staff)

Sara McLafferty (University of Illinois, Urbana-Champaign)

Jeremy Mennis (Temple University)

Wendy Nilsen (OBSSR)

Linda Pickle (StatNet)

Barbara Rapp (NLM)

Mike Ratcliffe (Census Bureau)

Jill Reedy (NCI)

Doug Richardson (AAG)

Lee Rivers Mobley (RTI International)

Elisabeth Root (University of Colorado, Boulder)

Gerry Rushton (University of Iowa)

Sheila Steffenson (Esri)

David Stinchcomb (Westat/NCI)

Daniel Sui (Ohio State University)

Zaria Tatalovich (NCI)

Daniel Wartenberg (Rutgers University)

John Wertman (AAG Staff)

Carolyn Williams (NIAID)

John Wilson (University of Southern California)

Li Zhu (NCI)

Appendix C
Workshop Agenda

Tuesday, February 22

8:30 – 8:40am Welcome and Introduction
Bob Croyle, NCI
Doug Richardson, AAG

8:40 – 9:00am Introduction of Workshop Participants

9:00 – 10:00am Panel 1: Overview of GIS Activities at Select NIH Institutes **Chair: Wilson Compton**
Wilson Compton, NIDA
Zaria Tatalovich, NCI and **Dave Stinchcomb**, Westat/NCI
Regina Bures, NICHD
David Balshaw, NIEHS
Carolyn Williams and **Rebecca Prevots**, NIAID

10:00 – 10:30am *Panel 1 Discussion*

10:30 – 10:45am BREAK

10:45 – 11:45am Panel 2: Perspectives from the GIScience research community **Chair: Mike Goodchild**
Mike Goodchild, University of California, Santa Barbara
Sara McLafferty, University of Illinois, Urbana-Champaign
Jonathan Mayer, University of Washington
Mei-Po Kwan, Ohio State University
Doug Richardson, AAG

11:45am – 12:15pm *Panel 2 Discussion*

12:15 – 1:15pm Lunch (on your own)

1:15 – 2:15pm Panel 3: Extramural Researcher Views: GIS and NIH **Chair: Jean McKendry**
Jarvis Chen, Harvard School of Public Health
Paul Gruenewald, Pacific Institute for Research and Evaluation
Gerry Rushton, University of Iowa
Elisabeth Root, University of Colorado, Boulder
Ellen Cromley, University of Connecticut School of Medicine

2:15 – 2:45pm *Panel 3 Discussion*

2:45 – 3:00pm BREAK

3:00 – 3:40pm Panel 4: System Architecture: Possibilities and Perspectives **Chair: Zaria Tatalovich**
Bill Davenhall, ESRI
Geoff Jacquez, BioMedware
Lee Rivers Mobley, RTI

3:40 – 4:00pm *Panel 4 Discussion*

4:00 – 5:30pm Breakout Groups
Common Needs – **Chairs: Bethany Deeds (NIDA), Sara McLafferty**
Key Challenges – **Chairs: Mei-Po Kwan, Jarvis Chen**
Implementation Alternatives – **Chairs: Dave Stinchcomb, Mike Goodchild**

5:30 – 5:45pm Wrap-up, Dinner Logistics

Wednesday, February 23

8:30 – 9:00am Summary of Breakout Groups from Previous Day **Group Chairs**

9:00 – 10:15am Discussion: Integration of Common Needs, Key Challenges **Group Chairs facilitate** and Implementation Alternatives

10:15 – 10:30am BREAK

10:30 – 11:40am Discussion: Recommendations, Priorities, and Next Steps **Chair: Doug Richardson**

11:45am – 12:25pm NIH Institute Leadership Executive Briefing: **Doug Richardson**
Summary of Key Outcomes and Discussion

12:25 – 12:30pm Closing Comments **Zaria Tatalovich**