Leadership in Biology
an AIBS initiative

Catalyzing advances in science, strengthening the profession, and bringing information to decision makers.

Addressing Biological Informatics Workforce Needs

A Report from the 2015 AIBS Council of Member Societies and Organizations Meeting

Meeting Sponsors

The Society for Integrative & Comparative Biology

Table of Contents

Executive Summary……3
Introduction…….4
Undergraduate Education……5
Graduate Training……...6
Training and Career Paths for Researchers and Practitioners…….9
Recommendations for Scientific Societies……12
Recommendations for Faculty Members...……13
Recommendations for Academic Libraries……...……13
Recommendations for Funders……...……...…13
Additional Resources…….14
Acknowledgments……..……14
Appendices…….……15

Executive Summary

The American Institute of Biological Sciences (AIBS) convened a meeting that brought together key stakeholders working to address education and training issues related to the biological informatics workforce. The meeting considered "biological informatics" broadly and defined it as the interdisciplinary science of collecting, curating, analyzing, publishing, planning, documenting, and archiving complex biological data, including molecular and sequence data.

Key discussions and proposed recommendations for future action are summarized in this report. Areas identified for continued development were undergraduate and graduate training, as well as the need for training and new career paths for informatics researchers and practitioners. The report offers recommendations for scientific societies, faculty members, academic libraries, and research funders.

Recommendations include:

1) Faculty should initiate training in essential elements of biological informatics, including the concept of the data life cycle, at the undergraduate level.
2) Project-based approaches and real data should be used in exercises when possible.
3) Students should understand core data skills, such as the fundamentals of data organization, principles of analysis and visualization, and basic concepts in statistics.
4) Professional societies should play a greater role in encouraging best practices and facilitating varied training experiences.
5) Professional societies should provide their members with data training opportunities.
6) Professional societies should encourage data sharing and open code through their journals' data publication policies.
7) Interdepartmental institutes (e.g., biology and computer science) might be a way to resolve the curricular tensions that arise between departments today.
8) Academic libraries should provide training on data sharing, the use of appropriate data, and metadata standards.
9) The community must continue to develop training materials and initiatives and increase its capacity to deliver training to graduate students, postdoctoral researchers, staff, and faculty.
10) Practicing scientists should take advantage of career training materials being developed by national organizations.
11) Postdoctoral fellowships in biological informatics are available, but more are needed.
12) National attention must be given to strengthening the workforce and the data infrastructure for biological informatics.

Introduction

The American Institute of Biological Sciences convened this workshop on Addressing Biological Informatics Workforce Needs in conjunction with its annual Council of Member Society and Organization meeting, which was held on December 8, 2015 in Washington, DC. This workshop included 12 invited speakers (Appendix 1) and additional meeting participants representing federal agencies, universities, scientific societies and research organizations, funders, and others (Appendix 2). Breakout groups synthesized key points from talks and developed recommendations for how various stakeholders can address pressing biological informatics workforce needs.

During this meeting and for the purposes of this report, "biological informatics" refers to the interdisciplinary science of collecting, curating, analyzing, publishing, and archiving complex biological data. The term "bioinformatics" is a close synonym but is often taken to refer to molecular data specifically; "biological informatics" encompasses molecular as well as other types of data, such as phenotypic, ecological, and specimen-based data.

This meeting built on the foundation laid by two previous AIBS workshops: Changing Practices in Data Publication and Enhancing Complex Data Integration across Research Domains. Reports from both workshops identified education, training, governance, and infrastructure as potential crosscutting barriers to the best use of an increasing amount and array of biological and environmental data. This plethora of data is a result of rapid technological progress, notably in instrumentation (including molecular techniques) and from digitization.

This report discusses the state of biological informatics training at undergraduate and graduate levels, as well as issues confronting midcareer professionals and individuals seeking to enter a

career path that has not yet been formally developed. In addition to identifying obstacles that, if ignored, could prevent biologists from gaining the most benefit from the rapid, ongoing changes in their science, this report identifies roles for scientific societies, journals, faculty members, research funders, and academic libraries.

Undergraduate Education

Importance

The importance of providing undergraduate students with exposure to data science was highlighted when participants noted that many doctoral programs rate students who have a bioinformatics background as better qualified than other applicants. Scripting, in particular, is a much-sought skill. Python and R are languages that provide a gentle learning curve for scripting and for learning how to handle data. Bioinformatics experience is marketable for those with undergraduate or graduate degrees, with R, Perl, Python, C/C++, Java, and MySQL being particularly sought-after skills.

According to workshop participants, there are too few high-quality programs providing undergraduate educational experiences. One example of an enabling resource is *CourseSource*, an online journal supported by the Howard Hughes Medical Institute, which allows people to upload classroom modules. The authors of these modules receive credit for a publication.

Departmental homes

Computer science departments can be good partners for biology departments that want their students to develop scripting skills, but achieving the necessary compromises between departments is sometimes an obstacle. An example shared during the meeting was of a scientific computing department that grew out of an interdisciplinary science program. However, the process was complicated by the challenges associated with aligning computer science and biology departmental objectives for curricular breadth and depth.

Continuing needs

Biologists often learn to code out of necessity. This model is not ideal and contributes to the development of bad practices. One participant argued that, in general, computer science departments are not good places for biologists to look for best practices in programming. The importance of improved instruction in statistics for all students was reaffirmed and offered as one part of the solution to preventing the adoption of bad practices.

DataONE, which has produced many informatics training materials, recommends that every undergraduate student taking an introductory level biology course would benefit from at least a one-hour lecture on how data are acquired and managed. Advanced undergraduates (along with graduate students and postdocs) who want to develop higher-level experience would benefit from taking a seminar course that covers best practices and tools for managing data throughout the data and research life cycles.

Undergraduate biology courses mostly do not require students to have data training, although there are exceptions. One is the Berkeley Data Science Education Program, which is starting to expose undergraduates to "computational thinking" through hands-on work with real data. However, informatics-training initiatives, which are often organized by professional societies, do not typically reach undergraduates, because they tend not to be much involved with professional societies. These programs, however, might offer an opportunity for professional societies to offer tangible, career-focused benefits to students.

The question of how to teach the teachers of data science is rarely addressed, and undergraduate education tends to rely on traditional courses and on faculty. Undergraduate data education also is often restricted to analysis and does not teach a broader understanding of data and data collection skills. Educators therefore need to define a core set of data skills that are desirable and marketable to employers; no consensus on that core set of skills exists at present.

The Quantitative Undergraduate Biology Education and Synthesis (QUBES) project has done important work centralizing quantitative course materials. However, faculty members can sometimes find it difficult to negotiate space in the curriculum for new material. Bringing research data into undergraduate classrooms has proven a promising way to teach data science (e.g., Teaching Issues and Experiments in Ecology, an initiative of the Ecological Society of America). Such projects could form the basis of a semester-long course or a workshop. Other promising approaches include small-group virtual meetings.

For systematics biology and the digitization of biodiversity collections, there is at the undergraduate level a need to develop cross-disciplinary integrations of data and a need to have organismal classes that incorporate taxonomy and field collections. There is also a need to teach data literacy and quantitative, geographic, and soft skills. Examples of soft skills are the ability to communicate effectively in written and oral formats, the ability to work independently and collaboratively, the ability to innovate, and the ability to make careful observations. QUBES, NIBLSE (the Network for Integrating Bioinformatics into Life Sciences Education), and AIM-UP! (Advancing Integration of Museums in Undergraduate Programs) have all done valuable work in these areas. Training in programming and the use of databases, as well as specialist tools such as GIS (geographic information systems), can even be delivered to high-school students in summer courses or labs. Teachers can retrieve modularized pedagogical material via the Internet and fit it into existing curricula as "microinsertions." Overall, educators need to better integrate the various efforts now under way. There is an unmet need for meaningful undergraduate research experiences involving specimen collection and curation, for mentoring of student workers, and for the use of educational modules that incorporate natural history collection data.

Graduate Training

Importance

Many thoughtful observers believe that society urgently needs more interdisciplinary research, as the tools for it now exist and solutions to many of society's most pressing problems span multiple disciplines. There remain many barriers to interdisciplinary research in academia,

however. Credit and incentives to encourage people to learn skills relevant to interdisciplinary research are not yet common. Perhaps partly in consequence, students with good data skills often leave academia for industry. The Gordon and Betty Moore Foundation's (GBMF) Data-Driven Discovery Initiative, a joint project with the Sloan Foundation, is one important example of a response; it will support researchers with computational, math, and statistics skills, as well as domain expertise.

Continuing needs

Reports starting in 2012, including from the Council of Graduate Schools, the National Research Council, the National Institutes of Health (NIH), and the American Chemical Society, have been critical of graduate student training broadly in the sciences. Principal criticisms included:

(1) The time needed to obtain a degree was too long.
(2) The master's degree was undervalued.
(3) The training was often narrow and provided few transferable skills.
(4) Career mentoring focused mainly on an envisaged future career in academe.
(5) Training was not aligned with disciplinary, workforce, societal, and student needs.

Recommendations for graduate training included:

(1) Expand/enhance professional skills.
(2) Prepare students for multiple career pathways.

(3) Create incentives for university/industry partnerships.
(4) More interdisciplinary training.
(5) Evidence-based approaches to increase retention and reduce time to degree.

Responding to these issues, the National Science Foundation (NSF) launched a new flagship graduate traineeship initiative—the NSF Research Traineeship (NRT) program. Established in 2014, the program is intended to (1) catalyze and advance cutting-edge interdisciplinary research in high-priority areas, (2) increase the capacity of graduate programs to produce interdisciplinary science, technology, engineering, and mathematics (STEM) professionals with technical and transferable professional skills for a range of careers, and (3) develop innovative approaches and knowledge that will promote transformative improvements in graduate education. Since its inception, the NRT has explicitly sought and funded new approaches to integrating data science into graduate education and to helping institutions build training capacity in data-enabled science and engineering. The program consists of two tracks:

- **The NRT Traineeship Track** is a traditional comprehensive, interdisciplinary graduate STEM traineeship in high-priority research areas. Funding is provided to institutions for up to five years with maximum awards of $3.0 million to support master's or doctoral degree students. The track has one priority interdisciplinary research theme—Data-Enabled Science and Engineering (DESE). One common feature of DESE awards is intensive, vertically integrated training, in which faculty and postdocs collaborate. This feature was introduced to counter the loss of expertise that previously occurred when students graduated.
- **The Innovations in Graduate Education Track** consists of smaller, three-year awards of up to $500,000. It provides no student support but, rather, pilots graduate education projects. These projects are intended to help students learn how to exploit data and use novel data-driven approaches.

From the experience of DataONE, it is recommended that graduate students take a seminar course that covers the best practices and tools for managing data throughout the data and research life cycles. Meeting participants supported exposing first-year graduate students to real data challenges. The Biodiversity Collections Network (BCoN) is building on efforts begun by AIM-UP! and working to foster the development of a community of practice that infuses specimen-based learning and exploration into formal and informal science education.

Students are often interested in data infrastructure and training but there are few programs available to which they can be directed. The lack of available training for graduate students in computing and informatics was noted by Hernandez and colleagues in *BioScience* (doi:10.1525/bio.2012.62.12.8) as a factor limiting data integration: Over 80 percent of students in California who participated in a survey said they had received no training in computing and informatics. A survey conducted by Strasser and Hampton attributed the deficiency to a shortage of time and to the lack of preparation of students and instructors. The National Center for Ecological Analysis and Synthesis (NCEAS) Distributed Graduate Seminars, which focus on a scientific synthesis project, are one promising route to providing such training. Another is NCEAS Summer Institute, which combines hands-on exercises and small group sessions. They have achieved success in teaching version control, data sharing, data "wrangling," and

collaboration skills. These institutes are, however, very over-subscribed. For example, in 2013, NCEAS received over 400 applications for 22 seats in one course.

Training and Career Paths for Researchers and Practitioners

Importance

According to an estimate by Change the Equation, 7.7 million people use complex computing in their jobs, which is 3.9 million more than the US Bureau of Labor Statistics reports.

"Data entropy" is a major problem in science: Data cited in older publications are very often unavailable. For publications more than 15 years old, more than half of the data referred to are no longer accessible, according to one estimate. Handling data constitutes a cycle that involves data production, data reuse, data cleaning, data exploration, and data preservation. Increasingly, we can expect to see automated processes involved. Yet when many researchers are asked what metadata standard they use, they answer "none" or "one created in my lab." The BCoN, as one example, seeks to enhance the training of existing collections staff and create the next generation of biodiversity information managers to improve the existing state of affairs.

Continuing needs

Early-career training is essential and should recognize that information scientists often do not have domain knowledge in the fields from which they are using data. Programming and the use of databases as well as specialist tools such as GIS can be incorporated into such training. Panelists recommended intensive "research sprints" or "hackathons" that bring together groups of researchers to solve problems quickly, particularly as a tool for teaching basic programing, data standards, and semantics.

Data-related intellectual property issues are important and not widely understood by researchers. For example, data cannot be copyrighted but specific data compilations can be. Training should therefore cover legal aspects of intellectual property, privacy issues, and similar concerns. Instruction in data management can conveniently and usefully be joined with ethics training, such as that mandated by the Responsible Conduct of Research provisions of the America COMPETES Act of 2007.

Published standards can and should continue to play an important role in helping researchers to adopt good data practices. Coursework in standards is an essential antidote to the widespread tendency to create a new data "standard" rather than use an existing one. This creates negative consequences: Data ostensibly supporting publications are being published in formats that make them impossible to reuse.

Formal semantic schemes are becoming increasingly important, not just for good data practice but for machine-learning approaches to text and data mining, for annotation, and for data integration. The OBOE ontologies for scientific observations, developed for the NCEAS, are one such valuable development.

Researchers must become comfortable with change and be willing to learn "on the job," as the tools for data intensive science are in a state of flux. They should be unafraid to ask questions, as a thriving online culture supports learners. GitHub and Stack Overflow are helpful sites for such learning.

Some commentators have argued that professional recognition for data skills is in general lacking in academia. A few universities are, however, making appointments in data science, and some federal agencies are setting up specialized data units, so it appears that attitudes are changing. Three data science environments supported by the Gordon and Betty Moore Foundation (at New York University, the University of Washington, and the University of California, Berkeley) have a Careers Working Group that seeks to find professional roles for computationally savvy researchers. The Foundation also supports the Jupyter interactive lab notebook for sharing workflows, the scientific programing language Julia, individual investigators, and Data Carpentry, an organization that runs scores of intensive advanced data training workshops around the country.

Data Carpentry recognized the demand for computational expertise several years ago. Many domain scientists have very little programming or computational experience when they start training. Data Carpentry is now helping to fill the strong unmet training need by providing domain-specific, hands-on intensive workshops around the country (34 events were held during 2015). These are developed by and for practitioners and aim to identify best skills and practices, with an emphasis on foundational skills. Their format as "add-ons" solves the problem of the lack of time in curricula for data training. Data Carpentry is also working on a "train the trainers" approach with Software Carpentry, a slightly older organization with some similarities. Several workshop participants voiced support for the sort of short workshops provided by Data Carpentry, as well as for cross-disciplinary hackathons.

A major effort at the NIH is the extramural Big Data to Knowledge (BD2K) initiative, which complements an intramural program. The BD2K expends funds in a disease-agnostic fashion on efforts that support data utilization across domains; data integration of a wide range of data types is a major focus. It aims to develop and improve data science skills, build a diverse workforce, ensure that training opportunities are available at all levels from undergraduate to senior faculty, and foster collaborations between data scientists and biomedical scientists. Almost 20 percent of the BD2K budget goes to training. This supports a wide variety of educational resources, courses (including massive open online courses), and training and career development programs. Funding is also available to give students at less research-intensive universities experience working with data science. The NIH is emphasizing making educational resources easy to discover through its training coordination center.

One meeting participant and observer of the "data deluge" in modern biology noted that the term "bioinformatics," although it might be taken to refer to any basic biological data, in practice is often understood as referring to molecular data. Biodiversity informatics, which can help researchers understand functional biodiversity, is very much a developing field. Although some understand it to concern analysis workflows and methods, others include within "biodiversity informatics" data infrastructure and knowledge assembly and provisioning. These aspects of biodiversity informatics have different audiences. The most fragile part is data infrastructure and training, which seem to be less valued than analytical methods and application services. There are fewer clear career pathways for those interested in the data side than for those on the analysis workflows side. Some incentives exist, but in general, there are too few opportunities on the data side.

A workshop held in September 2015 at NCEAS on data-intensive skills across the environmental sciences emphasized the importance of the concordance of the skill classes needed, such as data management and processing, software skills needed for science, analysis, visualization, communication and dissemination, and collaboration and synthesis. To scale up existing efforts, independent training should be encouraged, with workshops and materials coordinated through other organizations (e.g., Data Carpentry). Scale up will also depend on networked assessments to generate a higher-level view of what is needed. A networked graduate course run using Software Carpentry techniques at multiple universities could be part of the solution; it would focus on a different topic each year. An idea incubated at this meeting was based on the recognition that teaching can be very demanding. It may therefore be useful for less experienced educators, perhaps postdocs, to partner with more experienced instructors for a time, then move on to a different institution to train others once they have gained experience.

CyVerse (formerly iPlant Collaborative) has developed powerful computational resources that are now more widely available. CyVerse is dedicated to advancing team science and is recognized by the National Center for Biotechnology Information as a center for data provision. This designation makes it easier for individual researchers to use its services. CyVerse hews to a platform philosophy that recognizes the importance of working at scale and avoids the mistake of assuming that one size fits all. Users can deploy and use different "building blocks" as they need them in a cloud space called iPlant Atmosphere. They can also custom design appliances by extending existing components and publish their findings. In addition, CyVerse makes available a number of science APIs (application programming interfaces). It thus provides tools that allow

people to manage their digital assets and improve computational productivity. Over 40 herbaria are integrated into the collaborative, but usage is extending beyond plants and life sciences (e.g., breast cancer research, psychological and social research, and climate research). CyVerse has created a course on applied concepts in cyberinfrastructure and special topics workshops.

DataONE has provided informatics training in a range of venues and varying degrees of depth for the past six years, from screencast tutorials only a few minutes long to two-weeks-long intensive graduate courses. High-quality, modular slide presentations, handouts, and exercises enable faculty members to easily create and modify informatics seminars and lectures for students.

Recommendations

Breakout groups generated the following recommendations for various stakeholder groups on the basis of the discussions outlined in the previous sections.

Recommendations for Scientific Societies

Scientific societies should provide plenary sessions with keynote speakers, as well as sections on pedagogy and symposia exploring the importance of data curation. These sessions should be scheduled for well-attended portions of meetings and not limited to pre- or post-meeting workshops. Professional communities should also promote and enforce data archiving policies by encouraging their journals to publish articles on data science methodology and software.

Journal editors and researchers expressed concern that the federal push for data publication was requiring editors to take on an enforcement function (i.e., to ensure that authors publish adequate data), without financial support. A contrasting sentiment was that some journals profit from the products of research while not paying reviewers, so they are morally obliged to house data products that support research. Others wondered how many of the data associated with a research project will have to be made public and whether tangential material such as researchers' emails would be included. The community and funders must resolve these expectations and standards of practice.

The push toward publication of data has prompted biological societies and journals to encourage, if not require, the publication of original data supporting a scientific article with the report, possibly after an embargo period. Many journals require that data be made available to reviewers at the time an article is submitted, but some researchers object to this requirement. Shared standards of practice are required to prevent an uneven playing field and to ensure that data are indeed available.

Because reviewers are not trained adequately to evaluate the data submitted with a manuscript, many papers are now being published with data that have been processed, precluding replication of the original analysis or reuse of the data. This tendency undermines the value of data publication. Scientific societies could take responsibility for helping to prevent this damaging outcome.

Recommendations for Faculty Members

Universities should require basic data standards instruction for new hires as they now require human resources training. If developing a course between a biology department and another academic entity such as a library or information school proves impossible, creating a new interdisciplinary institute may be a solution. One participant shared the promising idea of having a computer scientist and a biologist co-teach a semester-long data science course at a university, with one postdoc partner. The postdoc would then move on to another institution to share the expertise.

Recommendations for Academic Libraries

Libraries were recognized as important potential transmitters of good data practices, and they are well positioned to help in the preparation of data management plans. Forward-looking libraries are already sharing information about basic data management and intellectual property awareness.

Recommendations for Funders

Funders, whether government or foundation, can play an important role in addressing education and training issues related to the biological informatics workforce. Individuals and institutions alike are prone to act when a funding source offers clear incentives. Additionally, funding sources should support the hiring of data professionals to advance their requirements.

Additional Resources

Ag Data Commons
https://data.nal.usda.gov

Advancing Integration of Museums in Undergraduate Education Research Coordination Network (AIM-UP!)
http://aimup.unm.edu

Berkeley Data Science Education Program
http://databears.berkeley.edu/program

Big Data To Knowledge (BD2K)
https://datascience.nih.gov/bd2k

Biodiversity collections Network (BCoN)
https://bcon.aibs.org

Course Source
http://www.coursesource.org

CyVerse (the iPlant Collaborative)
http://www.iplantcollaborative.org

Data Carpentry
http://www.datacarpentry.org

DataONE
https://www.dataone.org

Data-Driven Discovery Initiative
https://www.moore.org/programs/science/data-driven-discovery

Encyclopedia of Life's TraitBank
http://eol.org/info/traitbank

GitHub
https://github.com

NCEAS Distributed Graduate Seminars
https://www.nceas.ucsb.edu/research/dgs

NCEAS Summer Institute
https://www.nceas.ucsb.edu/outreach/summer-institute/2013/summer-institute-2013

National Science Foundation Research Traineeship (NRT) Program
http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf16503

Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE)
https://qubeshub.org/groups/niblse

OBOE Semantic Tools Project
https://semtools.ecoinformatics.org/oboe

Stack Overflow
http://stackoverflow.com

Teaching Issues and Experiments in Ecology
http://www.esa.org/tiee/misc/about.html

The Quantitative Undergraduate Biology Education and Synthesis (QUBES) project
https://qubeshub.org

Acknowledgments

AIBS gratefully acknowledges financial support for the meeting from the New Mexico Experimental Program to Stimulate Competitive Research (EPSCoR), the Society for the Study of Evolution, The Society for Integrative and Comparative Biology, the Ecological Society of America, iDigBio, the iPlant Collaborative (now CyVerse), the Biodiversity Collections Network, the Society for the Preservation of Natural History Collections, DataONE, the Computational Biology Institute of George Washington University, and the University of Kansas Biodiversity Institute.

The following AIBS staff members provided essential support for this meeting: Diane Bosnjak, Scott Glisson, Robert Gropp, Syreeta Jones, Jeffrey Newman, Julie Palakovich Carr, Jennifer Pettit, James Verdier, and Joel Wagener. Former AIBS staff member Tim Beardsley was instrumental in organizing this meeting and developing the report.

Appendix 1

Speaker Biographies

Timothy M. Beardsley

Timothy M. Beardsley was the Editorial Director of AIBS Publications and Editor in Chief of *BioScience*. He was a principal organizer of the NSF-supported workshop on Changing Practices in Data Publication that AIBS held in December 2014.

Keith A. Crandall

Keith A. Crandall is the founding Director of the Computational Biology Institute at George Washington University. Professor Crandall studies the computational biology, population genetics, and bioinformatics, developing and testing of big data methods of DNA sequence analysis. He applies such methods to the study of the evolution of both infectious diseases (especially HIV) and crustaceans (especially crayfish). Professor Crandall has published over 250 peer reviewed publications, as well as three books. He has been a Fulbright Visiting Scholar to Oxford University and an Allen Wilson Centre Sabbatical Fellow at the University of Auckland. Professor Crandall has received a number of awards for research and teaching, including the American Naturalist Society Young Investigator Award, an NSF CAREER Award, a PhRMA Foundation Faculty Development Award in Bioinformatics, Honors Professor of the Year award at Brigham Young University, ISI Highly Cited Researcher, and the Edward O. Wilson Naturalist Award. He was also recently elected a Fellow in the American Association for the Advancement of Science (AAAS). Professor Crandall earned his BA degree from Kalamazoo College in biology and mathematics, an MA degree from Washington University in statistics, and a PhD from Washington University in biology and biomedical sciences. He also served as a Peace Corps volunteer in Puyo, Ecuador. *kcrandall@gwu.edu*

Clifford Duke

Clifford Duke is the Director of the Ecological Society of America's (ESA) Office of Science Programs, which promotes the continued development of ecological science and its integration into decisionmaking and education. At ESA, he leads a wide range of projects, including outreach efforts for engagement of the US scientific community in the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES); Sustaining Biological Infrastructure, an annual short course on strategic financial planning for project directors; and support for ESA's Vegetation Classification Panel. Dr. Duke has led five workshops on data sharing issues, has contributed to the DataONE Community Engagement and Outreach Working Group, and currently serves on the National Research Council's Board on Research Data and Information. *csduke@esa.org*

Michelle Dunn

Michelle Dunn is Senior Advisor for Data Science Training, Diversity, and Outreach in the Office of the Associate Director for Data Science (ADDS) at the NIH. In the ADDS Office, Dr. Dunn's responsibilities focus on education, training, and workforce development in data science, as it is applied to the biomedical, behavioral, and clinical sciences. Prior to joining the NIH/OD, Dr. Dunn was a program director at the National Cancer Institute. She received her PhD in statistics from Carnegie Mellon University and her AB in applied mathematics from Harvard College. In her free time, she enjoys spending time with her family, whether visiting far-off places or baking at home. *dunnm3@od.nih.gov*

Robert Gropp

Robert Gropp is the Interim Co-Executive Director of the American Institute of Biological Sciences. He was a principal organizer of the NSF-supported workshop that AIBS held in May 2015 on Enhancing Complex Data Integration across Research Domains. Additionally, Gropp is the Principal Investigator on an NSF-funded Research Coordination Network grant that is investigating ways to build a sustainable national community capable of implementing the recommendations outlined in the "Strategic and Implementation Plans for a Network Integrated Biocollections Alliance." Gropp also serves as Director of Policy for AIBS and the Natural Science Collections Alliance. *rgropp@aibs.org*

Robert Guralnick

Robert Guralnick is Associate Curator of Biodiversity Informatics in the Department of Natural History, University of Florida, and the Florida Museum of Natural History. A biodiversity scientist, his research focuses on what causes spatiotemporal changes in genetic and species diversity. His lab takes an integrative approach to global change biology and the skills in the lab range from occupancy modeling to spatial ecological modeling, to landscape genetics, to molecular phylogenetics. Because so much of the work in his lab uses primary species and population occurrence data (when and where species and populations occur) available from natural history collections and citizen science naturalists, it is very involved in ecological and biodiversity informatics initiatives to increase the quality, availability, and utility of such data sets at the global scale. *rguralnick@flmnh.ufl.edu*

Stephanie Hampton

Stephanie Hampton is a Professor in the School of the Environment at Washington State University and Director of the Center for Environmental Research, Education, and Outreach. Her interests range from basic research in aquatic science using statistical analysis of large databases to broader applications of empirical evidence in environmental issues and policy. Her activities in education and training have shifted across a spectrum from traditional classroom instruction toward larger multi-institutional efforts, and are now focused on increasing the capacity of environmental researchers to engage in data-intensive science. With colleagues at various institutions she is coordinating and scaling up approaches to training for data-intensive skills across environmental sciences, with the support of an Emerging Frontiers grant from the NSF. *s.hampton@wsu.edu*

Nirav Merchant

Nirav Merchant is Co-Principal Investigator for CyVerse and director of BioComputing at the University of Arizona. His research involves addressing computational challenges in data life cycle management. The primary emphasis of his recent work has been on developing infrastructure and tools for enabling data-driven collaboration for geographically distributed teams that require high-throughput, large-scale data handling capabilities. His research interests include process automation and scalable and pervasive computing with applications in mHealth (Mobile Health) based interventions. He has served as key technology strategist for national and international projects, with over two decades of productive collaborations in the biomedical arena. *nirav@email.arizona.edu*

William Michener

William Michener is Professor and Director of e-Science Initiatives at the University of New Mexico's University Libraries and Project Director for New Mexico's NSF and Department of Energy EPSCoR Programs, and DataONE—a large DataNet project supported by the NSF. He is involved in research related to creating information technologies supporting data-intensive science, development of federated

data systems, and community engagement and education. He has a PhD in Biological Oceanography from the University of South Carolina and has published extensively in marine science, as well as the ecological and information sciences. He serves on several boards of nonprofit organizations and has expertise in project management and meeting facilitation. He is Editor of *Ecological Archives*, Associate Editor for *Ecological Informatics*, and a member of the Editorial Board for *Ecology*. *william.michener@gmail.com*

Anna Monfils

Anna Monfils is an Associate Professor and Director of the herbarium at Central Michigan University (CMU). Her background is in plant systematics, ecology, and evolution, and she has an active research program investigating systematics in *Fuireneae* (*Cyperaceae*), management of aquatic invasive macrophytes in the Great Lakes, and conservation in prairie fen wetlands. She mentors undergraduate, master's and PhD level graduate students. Dr. Monfils serves on the BCoN Advisory Council, is a participant in the AIM-UP!, was an iDigBio scholar, and has worked closely with iDigBio to develop the Small Collections Network (SCNet). She co-developed an NSF-URM Biology Undergraduate Mentoring Program, BUMP into Research at CMU!, and is a co-Principal Investigator on the PLANTS II grant to the Botanical Society of America. Both programs are designed to foster broader participation in the biological sciences. *monfi1ak@ccmich.edu*

Cynthia Parr

Cynthia Parr is a Technical Information Specialist in the Knowledge Services Division at the National Agricultural Library of the US Department of Agriculture (USDA). She leads the Ag Data Commons research data repository project. Previously she served as the Chief Scientist and Director for Species Pages for the Encyclopedia of Life (EOL), based at the National Museum of Natural History of the Smithsonian Institution. She currently serves as the Chair for the Biodiversity Information Standards organization (Taxonomic Database Working Group, TDWG). She serves on the USDA Enterprise Data Management working group, which is drafting departmental data policy and guidelines; is a member of the BioEco interdepartmental working group; and served on the 2016 Earth Observation Assessment Biodiversity Societal Benefit Area team. She has a PhD in biology from University of Michigan and has conducted research in evolutionary ecology, ornithology, behavior, molecular systematics, community ecology, information visualization, semantic web, and social networks. *cynthia.parr@ars.usda.gov*

Carly Strasser

Carly Strasser is a Program Officer at the Gordon and Betty Moore Foundation. She works within the Data-Driven Discovery Initiative, an effort within the foundation's Science Program focused on promoting both the researchers and the practices required for impactful data-driven research, with a broader vision of accelerating scientific discovery. Previously, Dr. Strasser was a Research Data Specialist at the California Digital Library. In this role, she was involved in development and implementation of many of the University of California Curation Center's services and worked to promote data sharing and good data management practices among researchers at the University of California and beyond. Her prior experience as a researcher in marine science and mathematical ecology has informed her work toward ushering in the new era of open, transparent, and collaborative science. Dr. Strasser received a BA in marine science and a PhD in biological oceanography from the MIT/WHOI Joint Program. *carly.strasser@moore.org*

Richard Tankersley

Richard Tankersley is Program Director in the Division of Graduate Education at the NSF, where he has responsibility for the NSF's NRT program and its predecessor, the Integrative Graduate Education and Research Traineeship (IGERT) Program. Dr. Tankersley is Professor in the Department of Biological Sciences at Florida Institute of Technology, where he studies the ecology, physiology, and behavior of marine invertebrates, particularly the processes mediating the recruitment, settlement, and metamorphosis of larval crustaceans. *rtankers@nsf.gov*

Tracy Teal

Tracy Teal is a Co-Founder and the Executive Director of Data Carpentry and an Adjunct Assistant Professor in BEACON at Michigan State University. She is on the NumFOCUS Board of Directors and on the Organizing Committee for the development of Reproducible Science Curriculum and the NCEAS Data Intensive Training initiative. Previously she was an Assistant Professor at Michigan State University in Microbiology and Molecular Genetics and an NSF Postdoctoral Fellow, where her research focused on bioinformatic and metagenomic approaches to understanding how agricultural practices affect soil microbial communities and their production and consumption of greenhouse gases. She has developed open-source tools for metagenomic, genomic, and image analyses and been a developer for Textpresso and Avida. As a proponent of enabling "big data" research, she is committed to training researchers in data management, analysis, and computational techniques. She is an instructor for the MBL STAMPS course and a co-founder and instructor of the EDAMAME course. As Data Carpentry Executive Director, she helps develop instructional strategies and open-source lesson materials for working with data in many domains, including ecology, genomics, geosciences, and social sciences. *tkteal@gmail.com*

Joseph Travis

Joseph Travis is Robert O. Lawton Distinguished Professor in the Department of Biological Science at Florida State University. He began his professional career in 1980 as an Assistant Professor at Florida State University, where he eventually rose to the position of Dean of the College of Arts and Sciences. Dr. Travis has served on numerous advisory panels and boards, including the Advisory Council for the Biological Sciences Directorate at the NSF. He is a Fellow of the American Academy for the Advancement of Sciences (1991) and was President of the American Society of Naturalists in 2005. He is currently President of the American Institute of Biological Sciences. His research focuses on the ecology of natural selection, including the effects of density-dependent selection on population regulation. In 2011, Travis was awarded the E. O. Wilson Naturalist Award from the American Society of Naturalists, which recognizes the contributions of midcareer scientists toward advancing the understanding of a particular ecosystem or group of organisms. *jtravis@bio.fsu.edu*

Appendix 2

Participant List

Tim Beardsley
American Institute of Biological Sciences
tbeardsley@aibs.org

Brett Burk
Burk and Associates, Inc.
bburk@burkinc.com

Janelle Burke
American Society of Plant Taxonomists
janelle.burke@howard.edu

Melissa Cragin
Directorate for Biological Sciences, NSF
mcragin@nsf.gov

Keith Crandall
Society of Systematic Biology
kcrandall@gwu.edu

Clifford Duke
Ecological Society of America
csduke@esa.org

Michelle Dunn
National Institutes of Health
dunnm3@od.nih.gov

Charles Fenster
Board of Directors, American Institute of Biological Science, and
University of Maryland
cfenster@umd.edu

Stephen Gallo
American Institute of Biological Sciences
sgallo@aibs.org

Scott Glisson
American Institute of Biological Sciences
sglisson@aibs.org

Robert Gropp
American Institute of Biological Sciences
rgropp@aibs.org

Robert Guralnick
University of Florida
rguralnick@flmnh.ufl.edu

Steven Hageman
Paleontological Society
hagemansj@appstate.edu

Stephanie Hampton
Washington State University
s.hampton@wsu.edu

Shelley James
iDigBio
sjames@flmnh.ufl.edu

Kiho Kim
International Society for Reef Studies
kiho@american.edu

Gary Krupnick
Association for Tropical Biology & Conservation
krupnick@si.edu

Hans Lemke
Association for Biology Laboratory Education
hlemke@umd.edu

Katherine McCarter
Ecological Society of America
ksm@esa.org

Marc Mendonca
Radiation Research Society
mmendonc@iupui.edu

Nirav Merchant
iPlant
nirav@email.arizona.edu

William Michener
DataONE and University of New Mexico
william.michener@gmail.com

George Middendorf
Herpetologists' League
gmiddendorf@howard.edu

Anna Monfils
Biodiversity Collections Network
monfi1ak@cmich.edu

Eric Nagy
Board of Directors, American Institute of Biological Sciences, and
Organization of Biological Field Stations
enagy@virginia.edu

Jeff Newman
American Institute of Biological Sciences
jnewman@aibs.org

Andrew Owen
George Mason University
aowen@aibs.org

Julie Palakovich Carr
American Institute of Biological Sciences
jpalakovichcarr@aibs.org

Jyotsna Pandey
American Institute of Biological Sciences
jpandey@aibs.org

Cynthia Parr
National Agricultural Library
cynthia.parr@ars.usda.gov

Deana Pennington
University of Texas at El Paso
ddpennington@utep.edu

Jennifer Petitt
American Institute of Biological Sciences
jpetitt@aibs.org

Arthur Shapiro
Department of Computer Science, American University
arthur.shapiro@american.edu

Judith Skog
Board of Directors, American Institute of Biological Sciences, and
George Mason University
jskog@gmu.edu

Jean Steiner
American Society of Agronomy
jean.steiner@ars.usda.gov

Carly Strasser
Gordon and Betty Moore Foundation
carlystrasser@gmail.com

Richard Tankersley
National Science Foundation
rtankers@nsf.gov

Tracy Teal
Data Carpentry
tkteal@gmail.com

Joseph Travis
President, American Institute of Biological Sciences, and
Florida State University
jtravis@bio.fsu.edu

James Verdier
American Institute of Biological Sciences
jverdier@aibs.org

Kathryn Walters-Conte
American University
kwalt@american.edu

Dennis Whigham
Society of Wetland Scientists
whighamd@si.du

Online Participants

Diane Bosnjak
American Institute of Biological Sciences
dbosnjak@aibs.org

Ian Johnson
American Institute of Biological Sciences
ijohnson@aibs.org

Jill Johnson
Society for Integrative & Comparative Biology
jill.johnson821@gmail.com

Josh Lancette
Entomological Society of America
jlancette@entsoc.org

Raymond Mejia
Society for Mathematical Biology
raymejia@gmail.com

Marilyn Norconk
American Society of Primatologists
mnorconk@kent.edu

Al Savitzky
Board of Directors, American Institute of Biological Sciences, American Society of Ichthyologists and Herpetologists, and Utah State University
savitzky@usu.edu

Joel Wagener
American Institute of Biological Sciences
jwagener@aibs.org

About AIBS

The American Institute of Biological Sciences is a nonprofit 501(c)(3) scientific association dedicated to advancing biological research and education for the welfare of society. AIBS works to ensure that the public, legislators, funders, and the community of biologists have access to and use information that will guide them in making informed decisions about matters that require biological knowledge. The organization does this through informing decisions by providing peer-reviewed or vetted information about the biology field and profession and by catalyzing action through building the capacity and the leadership of the community to address matters of common concern.

Founded in 1947 as a part of the National Academy of Sciences, AIBS became an independent, member-governed organization in the 1950s. Today, AIBS has individual members and nearly 160 member organizations with a combined individual membership of more than 200,000. AIBS is headquartered in Reston, VA, with a Public Policy Office in Washington, DC. The organization works to achieve its mission by publishing the peer-reviewed journal *BioScience* and the education website ActionBioscience.org, by providing scientific peer review and advisory services to government agencies and other clients, and by collaborating with scientific organizations to advance public policy, education, and the public understanding of science.

Membership Information

To become an individual or organizational member of AIBS, please visit http://access.aibs.org.

Social Media

Stay engaged by following AIBS on Twitter @AIBS_SPARS for news and information about peer review research and programs, @AIBS_Policy for science policy news and updates, and @BioScienceAIBS for the latest updates about upcoming articles and podcasts.

AIBS Member Organizations

(as of December 2015)

Academy of Natural Sciences of Drexel University
American Arachnological Society
American Fern Society
American Institute of Biological Sciences
American Malacological Society
American Museum of Natural History
American Ornithologists' Union
American Public Garden Association
American Society for Gravitational and Space Research
American Society for Photobiology
American Society of Ichthyologists and Herpetologists
American Society of Mammalogists
American Society of Naturalists
American Society of Parasitologists
American Society of Plant Biologists
American Society of Plant Taxonomists
American Society of Primatologists
Animal Behavior Society
Arizona State University
Arkansas State University
Association for Politics and the Life Sciences
Association for the Sciences of Limnology and Oceanography
Association for Tropical Biology and Conservation
Association of College & University Biology Educators
Association of Ecosystem Research Centers
Bailey-Matthews Shell Museum
Biodiversity Institute – University of Kansas
Biological Sciences Curriculum Study
BioQUEST Curriculum Consortium
Bishop Museum
Botanical Society of America
Brigham Young University
Brooklyn Botanic Garden
Cactus and Succulent Society of America
California Academy of Sciences
California Botanical Society
Coastal and Estuarine Research Federation
Coastal Education and Research Foundation
Columbus Technical College
Cooper Ornithological Society
Cornell University
Crustacean Society
Delaware Museum of Natural History
Denver Museum of Nature and Science
Ecological Society of America
Entomological Society of America
Florida Dept. of Ag & Consumer Services
Genetics Society of America
Harvard University-MCZ
Helminthological Society of Washington
Herpetologists' League
Hiram College
Human Anatomy & Physiology Society
iDigBio
Illinois State Museum
Instituto de Geologia Museo de Paleontologia
International Association for Bear Research and Management
International Association for Landscape Ecology, U.S. Division
International Society for Ecological Modelling
Kent Prior Parks Canada Agency
Lepidopterists' Society
Long Term Ecological Research Network
Louisiana State University-MNS
Michigan State University
Mississippi Museum of Natural Science
Missouri Botanical Garden
Mycological Society of America
National Association of Marine Laboratories
National Shellfisheries Association
National Tropical Botanical Garden
Natural Areas Association
Natural History Museum-LA County
Natural Science Collections Alliance
New York Botanical Garden
North Carolina Museum of Natural Science
Occidental College-MLZ
Organization of Biological Field Stations
Paleontological Society
Phycological Society of America
Poultry Science Association
Purdue University, Depart. of Entomology
Queens University of Charlotte
Radiation Research Society
San Diego Natural History Museum
Santa Barbara Museum of Natural History
Sitka Sound Science Center
Society for Behavioral Neuroendocrinology
Society for Conservation Biology
SDSU Natural Resource Management
Southern Appalachian Botanical Society
Southern California Academy of Sciences
Southwestern Association of Naturalists
St. Joseph's University, Biology Depart.
Texas A & M University
Texas Tech University
The Field Museum of Natural History
Torrey Botanical Society
Utah Museum of Natural History
University of Alaska, Museum of the North
University of Connecticut, Depart. of Ecology and Evolutionary Biology
University of Florida
University of Iowa, Museum of Natural History
University of Minnesota
University of Nebraska
University of New Mexico-MSB
University of Oklahoma
University of Texas-El Paso
Washington State University-NHM
Weed Science Society of America
Yale University

Suggested Citation

American Institute of Biological Sciences. 2015. *Addressing Biological Informatics Workforce Needs: A Report from the 2015 AIBS Council of Member Societies and Organizations Meeting*. Reston, Virginia.

Photo Credits

Julie Palakovich Carr, AIBS.