Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-05 06:00:11
REFERENCE NO: 221

Author Names & Affiliations

- Sarah Davidson - The Ohio State University, Max Planck Institute for Ornithology
- Gil Bohrer - The Ohio State University
- Roland Kays - North Carolina State University, NC Museum of Natural Sciences
- William McShea - Smithsonian Conservation Biology Institute
- Martin Wikelski - Max Planck Institute for Ornithology, University of Konstanz

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

biology, animal movement ecology

Title of Submission

Advanced cyberinfrastructure needs for the field of animal movement ecology

Abstract (maximum ~200 words).

The field of animal movement ecology has growing cyberinfrastructure needs to manage, archive, and analyze the explosion of data collected by animal-borne sensors and remote cameras. Community-wide repositories, data standards, and training opportunities are needed for quality control and assurance, data sharing and meta-analyses, interdisciplinary collaboration, shared analysis tool development, and to support data collection by individuals and organizations without the resources to maintain their own databases. Currently there are many short-term funding opportunities to discuss and design shared databases, but few if any financing methods to actually build and maintain such infrastructure. Due to a lack of community-wide data standards, the volume of unstandardized and poorly documented data is growing quickly, increasing the challenge of managing and archiving these data. Training opportunities are needed to ensure that researchers, students, and professionals have adequate expertise in GIS, database management, coding, data archiving, and remote sensing to thrive as movement ecologists.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

Data in the field of animal movement ecology is exploding in growth due to technological improvements of animal-borne sensors and remote cameras. As device size decreases and sensors improve the amount of data that can be collected has increased, allowing us to describe the animal movement process with orders of magnitude more detail than before. This provides exciting opportunities to pursue

answers to long-standing questions about animal ecology, moving from descriptive to hypothesis-based studies, often with applied conservation implications. However, such large data volume also generates new challenges in management, archiving, and analysis. Not only do these challenges limit the ability of researchers to conduct their own research, they also restrict cross-study comparisons due to the a lack of shared cyberinfrastructure with stable financing, standards for describing and providing data, and people trained to manage and analyze big data. Only when these technological, financial and human resources are in place will the discipline be able to reach a more mature stage where the data, theory, and application can advance in parallel. We've identified five research challenges in the field of movement ecology directly relating to cyberinfrastructure:

Long-term changes in movement patterns: Questions about how migration and other seasonal movement patterns are affected by environmental change, common to studies of impacts of changes in climate and land cover, require information on movements of many animals over an extended period of time. Currently this is best achieved by combining datasets from many collaborators; however the lack of data standardization and adequate funding for shared infrastructures make it much more difficult and expensive to do this.

Juvenile dispersal and long-distance movements of small animals: The size of even the smallest animal-borne sensors remains too large to safely deploy on most species or on juveniles of many more species. Juvenile dispersal is critical to gene flow between existing species, and for the establishment of new populations in newly suitable habitat in response to land use and climate changes. Understanding the rules followed by these dispersers would help connect critical habitat with appropriate movement corridors. Substantial decrease in battery and device tag size would vastly increase the number of species that can be studied using animal-borne sensors.

Comparative community ecology: Remote camera traps are unique in their ability to collect community-level data in a standardized way. Comparisons of these communities across sites and across years has led to important discoveries and data immediately relevant to conservation, but have so far been limited to the few sites sampled by larger organizations that support data infrastructure needed to manage camera tracking imagery.

Citizen science: Tracking tags and camera traps are sensors that could be deployed by citizens to collect reliable data over larger areas than scientists themselves can reach. However, proper cyberinfrastructure is needed to manage the data and support the citizen scientists.

Statistical tool development: Collaboration with statisticians and data miners offers movement ecologists opportunities to address analysis needs such as ways to visualize geospatial trajectory data and reliable methods for automated analysis of camera trap images. Many statisticians and data miners are interested in the opportunities and challenges associated with big data from animal movement, but need access to centralized repositories of curated and standardized datasets to develop new analytical tools.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Financing methods for sustaining cyberinfrastructure: As described above, shared infrastructure for collaboration and archiving is a necessity for research in movement ecology. However there are few if any ways to fund the maintenance and user support for such infrastructure, in particular for those that are shared beyond single agencies, universities or organizations. Ironically, there are many opportunities for short-term grants to discuss and perhaps prototype these infrastructure, but these fall far short of the financing needed to pay developers or to design a system with long-term use, growth and maintenance in mind. The result is many disconnected and poorly maintained systems with small groups of users and competition between funded projects. Researchers are often reluctant to participate in or contribute to shared infrastructure, concerned about research overlap or investing time in a system that will fall into disuse.

Data standards: The field of movement ecology is constrained by a lack of data standardization that makes interoperability, metaanalyses, and data archiving more difficult, expensive, and prone to error. Many researchers and manufacturers are not concerned with data archiving or standardization at the time data are collected and delivered, and because this has not been a priority, the number of devices and device manufacturers and the volume of existing unstandardized and poorly documented data are growing quickly. As new devices are being developed and modified in response to researcher needs, even a single research group often goes through multiple changes in manufacturer, device, and resulting data over the course of several years—combined with a lack of adequate data management expertise (more on this below), the consequence is that even the researchers who collected the data are often uncertain of basic information like the

time zone in which all their data were collected. The development of community-developed standards for documenting and transferring data would help provide a needed unified call for manufacturers to begin to consider data standardization an essential service for their customers.

Analysis tools: Development of analysis software that can be shared across the field require shared cyberinfrastructure to obtain sample datasets for development, understand data inputs, etc. As examples of analysis needs, for individual animal movements, while existing software provide sophisticated tools for displaying other types of geospatial information (raster and point data, 3D scenes, etc.), sophisticated and user-friendly methods for visualizing geospatial trajectories are comparatively few. In particular for large or high-resolution datasets, researchers in movement ecology regularly report difficulty in finding ways to visually explore their trajectory data. For remote camera data, methods are needed for automated animal identification; calculations of distance, size, and movement speed from camera trap images; and estimating populatin density using camera trap imagery and individual movement data.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

We see a critical need for training to help researchers, students, and professionals in the field of movement ecology to work with cyberinfrastructure and big data. Relevant skills include
• GIS concepts and software
• Database design and management
• Coding, e.g. SQL, R, MATLAB
• Data archiving concepts: data standardization, metadata, motivations for archiving
• Understanding of remote sensing data collection, processing, and distribution

While a segment of the movement ecology community is highly competent with many of these skills, it is still common to work with data stored in Excel files with no record of column definitions and data owners unclear of why or how information, such as timestamps, should be stored consistently within a database. Graduate students are increasingly exposed to coding, databases, and advanced analytical techniques, but this exposure varies widely and learning is typically focused on those skills most directly needed for their own research, rather than as important concepts needed to understand science and grow as researchers within the discipline.

To meet the cyberinfrastructure needs of movement ecology, these skills should become better integrated as part of curricula in departments that train movement ecologists. This might require training of current faculty or bringing in expertise from other departments to teach particular skills or courses. Undergraduate and graduate programs, as well as employers, should encourage and offer opportunities for development of these skills, either through specific courses or by making sure they are incorporated into other discipline-specific courses and continuing education.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”