Beyond ecoinformatics

The move toward synthesis and large collaborations in ecology requires more information technology than a lone scientist with a field notebook ever needed. As an integrative discipline asking increasingly broad and practical questions, ecology puts together very different kinds of information gathered in very different ways. Making those data work together is a huge challenge that opens the door to new discoveries and new ways of discovering. According to complexity scientist Eric Berlow, "The more you can zoom out and embrace complexity, the better chance you have of zooming in on the simple details that matter most."

Ecoinformatics has taken big strides in recent years, between data federations like DataONE bringing dispersed and heterogeneous ecological data under one searchable roof, and tools like R and Kepler tracking data transformations throughout an analysis. But data-intensive science is evolving rapidly, and not just within ecology. Here we survey the players in the data field, scanning the horizon for trends and tools that could inform how environmental science finds, stores, and uses data in the future.

This analysis includes organizations both inside and outside ecology. We survey projects and centers in ecoinformatics that are developing tools specifically for the study of ecology and biodiversity. Then we turn to informatics in disciplines that deal with other complex systems, like the weather and the human brain. Organizations doing research into informatics itself are covered briefly. Some of the most interesting advances in big data management and analytics are coming from the private sector—from multinational corporations handling databases on tens of millions of customers to a startup using network algorithms to detect disruptive trends in global technology. Another exciting space is data visualization, where tools are being developed to allow anyone, and sometimes scientists, to easily visualize data not just as an end product but as a means of data exploration (Fox and Hendler 2011). We list the data repositories holding the bulk of open access environmental data, plus Entrez, the life sciences search engine that is farthest along in terms of

moving seamlessly between data and literature. Finally, some attention is paid to data capture—from workflows that streamline manual data entry to sensor arrays that convert aspects of the physical world to digital form.

Each of the projects and organizations listed in the accompanying table and described in the accompanying document participate in one or more stages of the data lifecycle. The table indicates the main functions of each player:

- *standards* - developing or implementing standards that allow data integration and interoperability
- *management* - getting data into and out of databases, and tracking its origins and history
- *sharing* - allowing access to well curated data, whether from open repositories or between collaborators
- *analytics* - turning data into knowledge
- *visualization* - developing ways to explore data visually