Data: share and share alike

Clifford S Duke
Director of Science Programs, ESA

"The miracle is this – the more we share, the more we have." – Leonard Nimoy. That this aphorism applies to data will seem obvious to many scientists, presumably including Nimoy's archetypal Mr Spock. Sharing data is one of the most basic ethical tenets of science, but also one of the thorniest technological and cultural issues scientists face. This month's article by Crall *et al.* (pp 414–18) highlights the evolving need for ready access to data to inform problem-solving in ecology. Crall and colleagues focus on a specific issue, the need for greater knowledge about existing datasets to inform management of non-native and invasive species. Noting that more widespread data sharing could help identify and prevent potential invasions, they point out that it's not just the data; useful data sharing also demands greater attention to documenting metadata (data about data) and managing data quality from initial collection to publication.

Non-native species management is just one example of the need for increased data sharing. Other examples range from small-scale research on individual species to global-scale studies of the effects of climate change on ecosystem functions. Understanding emerging fields like biocomplexity requires synthesis of data from multiple sources and collaboration among scientists from a broad range of disciplines (see www.nsf.gov/geo/ere/ereweb/acere_synthesis_rpt.cfm). The development of large scale experimental systems exemplified by the Long Term Ecological Research (LTER) network and envisioned in the National Ecological Observatory Network also requires greater resources for data sharing. Virtually every field of ecology would benefit from more comprehensive and systematic documentation and publication of metadata, combined with general availability of data through online sources. Meeting these needs requires action by both organizations and individual scientists.

The Ecological Society of America (ESA) has been grappling with this issue since at least the mid-1990s. The 1995 Future of Long-term Ecological Data Committee report recommended greater collaboration among ecologists and other scientists to facilitate data exchange, and created the impetus for the launch of Ecological Archives, which publishes both data associated with journal articles and stand-alone data papers. Beginning in 1999, the Panel on Vegetation Classification developed VegBank, a publicly available vegetation plot database, and in 2004, the Ecological Visions Committee recommended that ESA create a data registry, which was launched this year (see data.esa.org). Also in 2004, the Science Office hosted the Society Summit, sponsored by NSF, which brought together the leaders of 12 major professional societies (including ESA) in ecology, evolution, and organismal biology, with the goal of developing common policies on data sharing and archiving. Three follow-on workshops are currently exploring common ground on, respectively, the development of data registries, requirements and recommendations for data centers, and identification and reduction of obstacles to data sharing.

Of course, ESA is not the only organization promoting data sharing in ecology or biology generally. A number of collaborative efforts described at www.ecoinformatics.org, including the Knowledge Network for Biocomplexity and the Science Environment for Ecological Knowledge, are contributing to the development of resources, tools, and standards for sharing data from a broad range of sources, with support from the National Center for Ecological Analysis and Synthesis, the LTER network, and university partners. Programs like the NSF's Cyberinfrastructure program, the Global Biodiversity Information Facility, the National Biological Information Infrastructure, and others are striving to provide resources to make data more discoverable and accessible.

The actions of professional societies, government agencies, and academic institutions are important; they provide the technology and the organizational and physical infrastructure that make data sharing possible. But technology and infrastructure are not the ultimate limiting factors for data sharing – the individual scientist is. For organizational efforts to succeed, individuals must commit to systematic changes in the way they work, from the planning of experiments to data collection through to publication and post-publication. Experiments must be planned from the beginning with documentation of data quality and collection of metadata in mind, not added as an afterthought. Authors need to register the supporting data in registries such as those sponsored by ESA, so that other researchers will know the characteristics of the data and where to find them. Researchers also need to work with their home institutions, professional societies, and funding sources to develop publicly available, permanent repositories to ensure that data are widely available and safeguarded against loss.

Such individual action is critical to solving emerging problems in ecology, vital to the health of the research enterprise, and a fundamental moral obligation on the part of scientists.