Harmanful nonnative species are spreading as the globalization of commerce increases the movement of terrestrial and aquatic organisms (Mack et al. 2000, Stohlgren et al. 2006). The inability to effectively combat these invasions has resulted in enormous environmental and economic losses worldwide (Pimentel et al. 2005). There are a number of individual efforts to integrate the available data, but a global cyberinfrastructure is required, one that incorporates a wide variety of data and makes that information available to anyone who is interested and has access to the Internet.

Finding and containing invaders early, while populations are small and concentrated, is essential for developing an early detection and rapid-response program to control invasive species (Rejmánek and Pitcairn 2002). To do that, field personnel need effective tools for identifying and recording invasive species, and for reporting them immediately to concerned agencies and individuals. The use of such tools would enhance programs in invasive species prevention, early detection, rapid assessment, rapid response, containment, and monitoring (Stohlgren and Schnase 2006).

Cyberinfrastructures are developing rapidly in the life sciences (Arzberger et al. 2004), including ecology (see www.neoninc.org) and the geophysical sciences (see www.geongrid.org). These cyberinfrastructures generally focus on making high-performance computers and large data sets available to scientists. A similar system could help prevent new invasive species from establishing and would make management of invasive species more efficient.

Requirements for a successful cyberinfrastructure
An effective invasive species cyberinfrastructure must (a) allow collection of data on the location and characteristics of invasive or potentially invasive species; (b) provide watch lists of potential new invaders by area; (c) send alerts for early detection of new invaders to concerned individuals; (d) model the current and predicted extent of an invasive species' range (Stohlgren and Schnase 2006); and (e) provide information on best-management practices for rapid responses to new invasions and for control and restoration efforts.

Documenting the spread of nonnative species is crucial; the species must be identified taxonomically, their characteristics noted, and their occurrence in space and time recorded. In addition, metadata must be added to this information to put it into context (when the data were collected, who collected the data, and how the data were processed). The minimum requirements for mapping a species are a location on

Jim Graham (e-mail: jim@nrel.colostate.edu) is a research scientist, and Greg Newman and Alycia Crall are research assistants, at the Natural Resource Ecology Laboratory at Colorado State University in Fort Collins; they study informatics for mapping the current and predicted ranges of invasive species. Annie Simpson is the invasive species coordinator for the National Biological Information Infrastructure at the US Geological Survey (USGS) in Reston, Virginia. Catherine Jarnevich is an ecologist, and Thomas J. Stohlgren is a branch chief, at the Fort Collins Science Center of the USGS in Fort Collins, Colorado. © 2008 American Institute of Biological Sciences.
Earth, a taxonomic identification, and a date of observation. These data can be used to create watch lists, send alerts, and construct models of species’ current and predicted future distributions.

Developing watch lists for effective prevention of invasions requires an understanding of the attributes of the invasive species in areas adjacent to the target area, a calculation of a habitat’s vulnerability to specific invaders inside the target area, and an extrapolation of the potential for spread from outside to inside the target area.

Early detection of nonnative species involves “smart surveys” (Stohlgren and Schnase 2006) of the most vulnerable habitats and the most invasive species to detect populations while they are small and can be affordably controlled or eradicated. When a new species is detected, a cyberinfrastructure can alert concerned people and organizations so they may begin eradication or management efforts for the infestation while it is small.

Modeling the current and potential range and abundance of an invasive species can be done with information on its existing and native ranges and abundances, including the physical parameters of the native and invaded ranges. Thus, location data for all organisms in both their native and introduced ranges are fundamental to predicting future invasions, and a range of environmental values is essential for modeling. Estimates of the vulnerability of a habitat to invasion—including margins of error—can be made on the basis of modeling results. Monitoring the location and abundance of an invasive species using predictive models, whether as part of long-term monitoring programs or to “ground truth” a predictive model (i.e., provide field data to validate the model), relies heavily on field personnel and on a stream of real-time location data. Examining the realized niche (or environmental envelope) of invaders relative to potential habitat for invasion can make data collection even more effective (Elith et al. 2006, Barnett et al. 2007).

Best-management practices for a rapid response to an invasive species, and for the most effective restoration efforts after the species is removed, can vary depending on the climate and the resources available for treatment. To develop best-management practices for a region, an efficient system that allows information exchange among port inspectors, field personnel, modelers, and resource managers needs to be in place for coordination of activities at various scales (Ricciardi et al. 2000).

**Architecture**

The proposed cyberinfrastructure would comprise a set of computers (“servers”), including databases and other software, that communicate through the Internet using Web service protocols. The cyberinfrastructure would eventually contain hundreds of servers, at a variety of organizations, that communicate and share data with each other through the Internet. Other computers, referred to as clients, would access the cyberinfrastructure through standard Internet browsers and special data collection software. These computers would include (a) programs for collecting data on the locations of invasive species, (b) databases to hold this information, (c) Web servers with remotely sensed data for modeling species distributions, (d) high-performance computers for modeling, (e) Web sites that provide access to the data through standard Web browsers, and (f) Web services to facilitate data exchange (figure 1).

![Figure 1. Architecture for an invasive species cyberinfrastructure.](image)

A great deal of work has already been completed, and more is ongoing, to create just this type of cyberinfrastructure. There are a number of efforts specifically targeted at invasive species, and we discuss these in the sections below. There is also a growing number of resources that were not created specifically for an invasive species cyberinfrastructure but can contribute to it. A work in progress, the Encyclopedia of Life (www.eol.org) will provide Web pages that describe millions of species across the globe, and invasive species Web sites can link to this site to obtain information about specific invasive species. The Global Biodiversity Information Facility (GBIF; www.gbif.org) contains more than 135 million species occurrence points that can be used for mapping and modeling invasive species. The Ocean Biogeographic Information System contains similar information; it has more than 16 million records for marine plants and animals (www.iobis.org; Halpin et al. 2006). Morphbank is a repository of photographs of species that can be used for species identification (www.morphbank.net). Discover Life (www.discoverlife.org) has information that can help identify species, including many that are invasive, as well as maps of species occurrences and photographs. The TeraGrid program (www.teragrid.org) provides access to high-performance computers for scientific research. The National Aeronautics and Space Administration (NASA; www.nasa.gov) provides access to large repositories of remotely sensed data from satellites and airborne sensors. Remotely sensed information is also available at a large number of commercial Web sites, such as Google Earth (http://earth.google.com), DigitalGlobe (www.digitalglobe.com), and i-cubed (www.i3.com). These are just a sample of the quickly developing Internet resources that could aid an invasive species cyberinfrastructure.
**Data collection tools**

Periodic examinations of areas to record new invasive species and to monitor the abundance and distribution of others are critical to predicting the rate of range expansion and the effectiveness of controls. Such field surveys can trigger alerts when new species are found, and subsequent surveys can provide ground-truth data to validate the accuracy of predictive models.

A significant amount of field data for invasive species surveys is now collected on paper forms. The geographic position of an organism is captured with a global positioning system (GPS) and recorded on the form, and the forms are filed or the data are entered into a spreadsheet or database after the field survey is completed. This process is often tedious, requiring many hours of data entry, and it is susceptible to data-entry error.

Digital devices can streamline the collection process. Devices such as personal digital assistants (PDAs) can be equipped with GPS and software designed for the collection of environmental data, allowing users to collect data and automatically upload them from the PDA onto a computer (WIMS 2006, Graham J et al. 2007). This process saves time and money, and reduces the potential for error by eliminating human transcription.

With the development of wireless fidelity systems and satellite communication, field tools can be linked directly to the Internet; thus, PDAs can submit data directly from field crews to the invasive species cyberinfrastructure. In addition, cyberinfrastructure handheld tools will be able to download maps and identification guides on the current location of the user.

**Databases.** There are many databases that contain data on current and potential invasive species, but only a fraction of these databases are available online (Crall et al. 2006). The Global Invasive Species Information Network (GISIN; www.gisinetwork.org) maintains a list of more than 200 online information systems that have a large variety of documentation on invasive species, including the current and past locations of invasive species infestations, species characteristics (e.g., life history), and the status of species in a given area (i.e., area of infestation). Some also provide species checklists, bibliographic references, image galleries, distribution maps, management plans, or identification guides (www.discoverlife.org; www.issg.org/database; Graham J et al. 2007).

The GBIF, which provides access to more than 135 million species occurrence records from more than 600 online databases, is the largest single repository for such data. The focus of the GBIF is on museum and herbarium collections, so only about half of the data contain location coordinates. The GBIF can be a valuable source of information for modeling invasive species, but the specific needs of the invasive species community also must be addressed.

There are a number of online databases for invasive species occurrence data, such as the Southwest Exotic Plant Mapping Program (www.invasiveweeds.com/mapping). Most of these databases focus on a specific region or taxonomic group. With some online systems, such as the Global Organism Detection and Monitoring system, users can document an organism’s occurrence and enter its associated data by clicking on a map or by uploading data sets composed of columns of data (text files and Shapefiles). This allows users to enter data from existing maps and paper forms (Graham J et al. 2007).

The Invasive Species Specialist Group maintains the Global Invasive Species Database (GISD; www.issg.org/database), which contains extensive information for more than 400 species, including taxonomy, life history, and management options. The GISD is mirrored by the US National Biological Information Infrastructure, which hosts a number of biological Web sites and invasive species databases (www.nbii.org).

It is currently very difficult to obtain a large collection of data with the locations of nonnative species (Crosier 2004). The proposed cyberinfrastructure would allow existing databases to be connected so that users could access and consolidate information easily and quickly. The cyberinfrastructure would need to integrate the latest information on nonnative species location data, temporal data, organismal attributes, and auxiliary data on environmental characteristics. The cyberinfrastructure would also need to provide data access through different user interfaces.

**Web sites.** Web sites can allow users to access the data and modeling capabilities of the cyberinfrastructure through standard Internet browsers. The invasive species cyberinfrastructure can contain a large variety of Web sites that target different groups of users and different regions of the world.

Resource management Web sites would give resource managers customized information and tools to help manage invasive species in their area. The National Institute of Invasive Species Science (NIISS) hosts a Web site (www.niiss.org) that helps resource managers control invasive species by using field tools to capture information, mapping species distributions, statistically analyzing data, modeling current and predicted distributions, and formulating the most effective control strategies (Graham J et al. 2007).

Citizen-science Web sites can provide access to models and invasive species data, as well as expose citizens to the scientific process (Ellisman 2005). Studies that engage citizen scientists are more likely to collect data relevant to local conservation and management issues (Danielsen et al. 2005). The Invasive Plant Atlas of New England (http://nbii-nin.ciesin.columbia.edu/ipane) is one example of a Web site that allows citizens to view invasive species information; after training, those citizens may submit data to a growing regional database.

Another benefit of an invasive species cyberinfrastructure is that e-mail alerts can be sent to concerned individuals when invasive species have been detected. The Nonindigenous Aquatic Species Web site (http://nas.er.usgs.gov) e-mails subscribers when new species arrive.
Data from the invasive species cyberinfrastructure would be used in conjunction with other cyberinfrastructure data. For example, a resource manager may create models to predict fire risk by overlaying previous fire location and intensity data from LANDFIRE (2007) on a map of the current distribution of invasive species. Another resource manager who is actively fighting fires may view information on how to keep certain invasive species from spreading into recently burned areas on a site dedicated to fire suppression.

The problem is that each Web site typically communicates with only a single database, greatly limiting the ability of users to see data outside a single taxonomic group or area. The Nonindigenous Species Database Network (NISbase; www.nisbase.org) brings together nine different invasive species databases into a single Web site (Simpson et al. 2006). The NISBase is the first step toward integrating individual invasive species databases, but to ensure that all the databases and Web sites can be integrated, a standard protocol for invasive species data is required (Simpson 2004).

Remotely sensed data. There is a wide range of environmental information that could be used for modeling invasive species abundance and distributions (Rowlinson et al. 1999, Everitt et al. 2002, Morisette et al. 2006). These data include Web mapping service (WMS; www.opengeospatial.org/standards/wms) sites such as NASA’s Jet Propulsion Laboratory (JPL; www.jpl.nasa.gov), which provides remotely sensed data and digital elevation models. The WMS protocol allows Web servers to call the JPL site and request portions of remotely sensed data at various resolutions and for different areas. The National Digital Forecast Database (www.weather.gov/ndfd) uses Web services to provide weather information for US locations. Systems such as the National Ecological Observatory Network will provide additional Web services for accessing environmental information.

Some invasive species can be detected from remotely sensed data (Everitt and Deloach 1990, Rowlinson et al. 1999, Everitt et al. 2002). A cyberinfrastructure for invasive species could acquire remotely sensed data, process them to find invasive species, and then make the data available on the Internet.

There is little that needs to be done to integrate remotely sensed data into modeling for invasive species other than working with agencies such as NASA to ensure that the required data continue to be provided through Web services.

High-performance computers. Species range models can be created on a typical computer workstation, but as the number of species occurrences and the spatial extent of the examined area grow, computer performance must improve. Eventually, high-performance computers will be required to produce results as quickly as Internet users expect.

Grid systems connect a large number of computers to create a virtual supercomputer. These systems, including TeraGrid, are being developed to perform beyond the capabilities of even the largest existing supercomputers (Reed 2003). It will be necessary to ensure that the models selected can be deployed on grid computer systems so the invasive species cyberinfrastructure can take advantage of these existing systems, and potentially contribute additional computers to the grid.

Web services for data exchange. Web services allow servers and clients to exchange data on the Internet through standardized communication protocols. Web services are typically enabled on a server by installing software that translates the data provider’s database into a standard protocol. These services also map the large number of formats for invasive species data into a single form, making it much easier to access the data. Web services then become the glue that holds together the databases, field collection tools, Web sites, remotely sensed data servers, and high-performance computers to create the invasive species cyberinfrastructure.

Web services are commonly used to provide news, advertising, and real-time weather reports on the Internet. Really Simple Syndication is a protocol used by Web services to communicate news headlines throughout the Web. Google and other advertising Web sites allow users to include a link to an advertising Web service on their own sites.

Georeferenced raster data, including remotely sensed data, are widely distributed through the Internet using the WMS protocol. These data are some of the most challenging kinds of information to move through networks because of their large size and because they cannot be compressed very much without losing information crucial for analysis and modeling. However, WMS has proved to be robust and has performed well.

There are a variety of existing standard protocols for the exchange of biological data. Two such protocols are the Distributed Generic Information Retrieval protocol (DiGIR; http://digir.sourceforge.net) and Biological Collection Access Services (BioCASE; www.biocase.org). The Taxonomic Database Working Group (TDWG; www.tdwg.org), an international standards group affiliated with the International Union of Biological Science and now known as Biodiversity Information Standards, is proposing the TDWG Access Protocol for Information Retrieval (TAPIR) as a mechanism to integrate data from various protocols, including DiGIR and BioCASE.

One problematic issue with Web service protocols is their performance when under high levels of utilization. When a protocol is new, it is used by only a few users with, typically, relatively small data sets. As a protocol becomes more popular, the number of users requesting data from the same server may surge. If the server does not respond quickly, either because of a large request or because it is busy servicing other requests, a user may think the service has failed, or that the browser has timed out, in which case the user receives a “Web site not available” message (also known as a “denial of service” error). The same problem can occur when servers request information from other servers. Another issue with Web services is that modeling systems may require thousands to millions of points. These problems combine to indicate that Web services for invasive species data will be
expected to run near the high level of performance that today’s systems are capable of. For example, the JPL Web service, which transfers raster images using the WMS protocol, can move large data sets at speeds of over 250 kilobytes per second between computers on the Internet (Graham 2006).

Tests have shown that the time to obtain records through the DiGIR protocol can vary as much as 100-fold, depending on the provider software configuration and target database. Extracting data for one species from an existing provider’s database can take as long as 15 hours. The time depends primarily on the way the Web service software interacts with the database, but the complexity of the protocol and the Web service software can be significant factors (Graham JJ et al. 2007).

One approach to improving cyberinfrastructure performance is to have a single server periodically harvest the data from each provider, store them as cached material in a system with optimized performance, and have users request data from the single server. This approach both improves performance and reduces the reliability of a distributed system. If the performance problems are not addressed, it could take years for the single server to update itself with the latest data from all providers. Adding hardware could improve performance, but would increase the costs that already strained research budgets must bear. The best long-term answer is to ensure that the protocols and software achieve the highest performance levels that the hardware is capable of.

None of the existing Web service protocols provides the type of information required to exchange data specifically for invasive species. These data include information on where a species is causing harm, characteristics describing the health of the organisms, and the management efforts that are being undertaken to control the species (Simpson et al. 2007). To facilitate computer-based information exchange on invasive species, a GISIN group operating simultaneously as a task group within TDWG has created a protocol to specifically address the needs of invasive species data exchange. This invasive species protocol is built on the TAPIR protocol. The protocol cannot support all of the details available at each invasive species Web site—the protocol would be too complex, and the funds available for its development and for the software to support it are limited. Instead, the protocol will support the needs that are shared across the databases and users of invasive species data. The protocol will also allow users to link to the original source of the data to obtain more information from a specific database. The protocol can also be expanded in the future as additional common requirements are identified. This protocol can be added to a data provider’s Web site by installing a toolkit available at www.niiss.org/GISIN.

Benefits

The immediate benefit of a cyberinfrastructure for invasive species would be to allow Web users to see data from different databases on one Web site, as opposed to having to find and visit a series of Web sites. For example, data on the zebra mussel (Dreissena polymorpha), which is invasive in the Great Lakes region and elsewhere in the United States, can be found at a large number of Web sites, each one of which would have to be visited and searched individually to amass all of the available information. Web portals such as NISbase can search across databases, thereby greatly increasing a user’s efficiency. With a broadly adopted standard protocol, Web portals could access a much larger set of databases with the same level of funding, which would also help with the development of different types of Web portals for different users (Maurer et al. 2000).

Members of the genus Tamarix (also known as salt cedar) are invasive in the southwestern United States. NASA, working with other members of NIISS, created a potential distribution of Tamarix for the United States by painstakingly combining field data from more than 45 databases, correlating them with remotely sensed data, and generating a predicted distribution map that can be used to prioritize management efforts (Morissette et al. 2006). The cyberinfrastructure described in this article could automate the entire process and make the distribution map available online so anyone could repeat the process for any species.

Efforts are under way to inspect vehicles and containers at key entry points to prevent the spread of invasive species (McCullough et al. 2006). The cyberinfrastructure could provide inspectors with prioritized lists of invasive and potentially invasive species that are likely to turn up, considering the origin of the container and its cargo. The inspectors could also add data to the cyberinfrastructure that would alert resource managers to be on the lookout for invasive species that were recently intercepted in shipments.

Conclusions

The challenges for the invasive species cyberinfrastructure are many. The maintenance and support for an invasive species cyberinfrastructure will require long-term, stable funding (Simpson et al. 2006). Most of the components of the cyberinfrastructure that are now in development are funded with relatively short-term research grants. Possible long-term funding sources include grants from foundations, additional import taxes, and fees for use of the cyberinfrastructure system.

Another challenge is to encourage researchers to make their data freely available, while ensuring continued credit for the original work (Maurer et al. 2000). Within the invasive species community, most researchers recognize the need for and the urgency of making data available, and most of the systems mentioned in this article allow data to be republished if credit is given to the original source. A record of the original authors of the data can be maintained by an electronic system, but it will always remain the responsibility of those using the data to credit the original source.

The cyberinfrastructure will work better for some invasive species than for others. Some invasive species can be detected from aerial surveys, but others cannot. It is easy to model
the ecological niche of some species, but some species are generalists and cannot be modeled effectively. In the end, the cyberinfrastructure will significantly improve the way we share data, create models, and communicate information on invasive species, but it will only be one suite of tools among many.

The quality of the data within the cyberinfrastructure will vary. Data collected by professional researchers tend to have better taxonomic identification than those collected by non-professionals. The use of PDAs to send data directly to a database, including photographs for species verification, can improve taxonomic accuracy. Much of the data in herbariums and museums lack the spatial information required for modeling. Because the level of quality required by different uses of the data also varies, the cyberinfrastructure should encourage and maintain the accuracy of all data elements so that users can make informed decisions about which data are appropriate to use for each application.

To make biological cyberinfrastructures successful, we need to ensure that (a) the benefits of the cyberinfrastructure are effectively communicated to ensure the visibility required to obtain long-term funding; (b) the GISIN protocol is developed within the required features, performance, and reliability; (c) relationships with remotely sensed data providers are maintained; and (d) relationships with high-performance computing services are developed. If we meet these needs, we will have created an international resource that will aid in the management of invasive species and thereby benefit everyone.

Acknowledgments

This research was funded by a National Science Foundation grant (OCI-0636213); a NASA grant (NRA-03-OES-03); the USGS Invasive Species Program; the National Biological Information Infrastructure; and the Colorado Agricultural Experiment Station. Logistical support was provided by the USGS Fort Collins Science Center and the Natural Resources Ecology Laboratory at Colorado State University.

References cited


doi:10.1641/B580312
Include this information when citing this material.