ECOLOGICAL INFORMATICS 2 (2007) 177–183

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/ecolinf

Full Length Articles

A global organism detection and monitoring system for non-native species

Jim Graham[a,*], *Greg Newman*[a], *Catherine Jarnevich*[b], *Rick Shory*[a], *Thomas J. Stohlgren*[b]

[a] *Natural Resource Ecology Laboratory (NREL) at Colorado State University (CSU), Fort Collins, CO 80523, USA*
[b] *USGS Fort Collins Science Center, Fort Collins, CO 80526, USA*

ARTICLE INFO

Article history:
Received 1 September 2006
Received in revised form
8 March 2007
Accepted 11 March 2007

Keywords:
Invasive
Database
Internet
Non-native
Detection
Monitoring

ABSTRACT

Harmful invasive non-native species are a significant threat to native species and ecosystems, and the costs associated with non-native species in the United States is estimated at over $120 Billion/year. While some local or regional databases exist for some taxonomic groups, there are no effective geographic databases designed to detect and monitor all species of non-native plants, animals, and pathogens. We developed a web-based solution called the Global Organism Detection and Monitoring (GODM) system to provide real-time data from a broad spectrum of users on the distribution and abundance of non-native species, including attributes of their habitats for predictive spatial modeling of current and potential distributions. The four major subsystems of GODM provide dynamic links between the organism data, web pages, spatial data, and modeling capabilities. The core survey database tables for recording invasive species survey data are organized into three categories: "Where, Who & When, and What." Organisms are identified with Taxonomic Serial Numbers from the Integrated Taxonomic Information System. To allow users to immediately see a map of their data combined with other user's data, a custom geographic information system (GIS) Internet solution was required. The GIS solution provides an unprecedented level of flexibility in database access, allowing users to display maps of invasive species distributions or abundances based on various criteria including taxonomic classification (i.e., phylum or division, order, class, family, genus, species, subspecies, and variety), a specific project, a range of dates, and a range of attributes (percent cover, age, height, sex, weight). This is a significant paradigm shift from "map servers" to true Internet-based GIS solutions. The remainder of the system was created with a mix of commercial products, open source software, and custom software. Custom GIS libraries were created where required for processing large datasets, accessing the operating system, and to use existing libraries in C++, R, and other languages to develop the tools to track harmful species in space and time. The GODM database and system are crucial for early detection and rapid containment of invasive species.

* *Corresponding author.*
E-mail addresses: jim@nrel.colostate.edu (J. Graham), newmang@nrel.colostate.edu (G. Newman), catherine_jarnevich@USGS.gov (C. Jarnevich), rshory@nrel.colostate.edu (R. Shory), tom_stohlgren@USGS.gov (T.J. Stohlgren).

1574-9541/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2007.03.006

1. Introduction

Non-native species cost the United States an estimated $120 billion per year in control and eradication programs and in reduced agricultural productivity (Pimentel et al., 2005). In addition, invasive species are viewed as the second greatest cause of decline in species diversity after habitat destruction (Wilcove et al., 1998). Invasive diseases impact human health and are a global problem (Mack et al., 2000), as evidenced by frequent news reports of Asian bird flu, West Nile Virus, or plague. However, information on the distribution of non-native plants, animals, and pathogens is largely held in widely disparate formats ranging from taxa-specific or region-specific databases to paper documents (e.g., herbarium specimens, range maps, field data sheets). According to a recent survey by the H. J. Heinz Foundation, there are at least 319 databases containing data on invasive species distributions in the United States alone (Crall et al., 2006). The information in these databases is rarely shared between organizations, and even more rarely formatted in consistent ways to facilitate sharing.

Largely due to global trade and transportation, harmful non-native species continue to invade (and re-invade) many areas. States, provinces, agencies, and non-government organizations are challenged to expend resources wisely to remove or contain invasive species to protect native species and ecosystem services (Mack et al., 2000). Based on global efforts to cooperate on invasive species issues (e.g., the Global Invasive Species Information Network) and interviews with resource managers in the central and eastern United States, it was clear that a dedicated system was needed to detect, map, and model harmful invaders to help combat invasive species.

Most of the information on organism distributions available on web sites provide annually updated regional information such as the Southwest Exotic Mapping Program, Nonindigenous Aquatic Species database, and the US Department of Agriculture's PLANTS database. Many web sites also provide the ability to download existing data including the Southern Appalachian Information Node of the National Biological Information Infrastructure, and the Invasive Plant Atlas of New England. Few web sites have the sophistication of VegBank (www.vegbank.org), created by the Ecological Society of America's Panel on Vegetation Classification. VegBank contains species information on over 30,000 vegetation plots located across the United States. This web site allows users to upload plot information, combine it with other plot data, and download combined results. It also allows users to browse plots by region or project and examine detailed plot data, including the percent cover of species. The Global Learning and Observations to Benefit the Environment (GLOBE) web site is an interagency program funded by the National Aeronautics and Space Administration (NASA) and the National Science Foundation (NSF) that focuses on teaching students how to take weather and biological measurements. The GLOBE web site allows users to enter scientific information on-line, view maps of where information is available, and download information the user has uploaded along with other user's information. The maps are rudimentary and there is no provision for the identification of native or invasive species, and thus no link to control information or long-term monitoring capability. Finally, the Distributed Generic Information Retrieval (DiGIR) system, developed by the Biodiversity Research Center at the University of Kansas, allows users to query biological museum records through a standard interface. Museums are provided the software to place their collection online through the DiGIR web site. DiGIR is used by many museums to provide access to their databases, but it was not created to allow users to update the data over the Internet (DiGIR, 2007).

With emerging technologies, we are now able to build global biological databases that are accessible to multiple scientists, resource managers, policymakers, and the public for a variety of purposes (Bowker, 2000). The most important factors in the success of large-scale biological information systems are how they handle the complexities arising from the precise locations and coexistence of millions of species, the relationships between species and environmental elements, and the communication and coordination required to work with various human organizations (Schnase et al., 2003). Our Global Organism Detection and Monitoring (GODM) system solves these problems by focusing on a simplified representation of biotic and abiotic information emphasizing the control of invasive species using technology that is available to users via the Internet. This advanced system was designed specifically to meet the needs of resource managers, agencies, and non-governmental organizations that manage invasive species on a daily basis. Other target users include 'citizen scientists,' researchers, decision makers, and the public.

Interviews with over 20 resource managers showed that the system should allow users to enter data on the spatial distribution of invasive species from spreadsheets, from Earth Science Research Institute (ESRI) Shapefiles, from geographic positioning systems (GPS), by directly typing coordinates on-line, and by clicking on on-screen views of standard topographic maps. The most requested feature was the ability to obtain a map of data that had been entered. Local resource managers and county weed coordinators wanted this map to be printable and available to the public over the Internet. Other important features include information on how to control and mitigate invasive species, "watch lists" of possible arriving species, and notification when new species were observed in their area(s) of interest. The ability to download data compiled by the web site from a variety of sources and incorporation of the data into standard reports was also desired. Additional beneficial features would include maps displaying the predicted spread of invasive species and analysis tools to determine the most effective control strategies in different environments. It also became clear that the system needed to be very easy to use with minimal computer expertise and had to work over standard phone-line modems, as many of the land management offices do not have high-speed Internet access, especially in remote areas and developing countries.

In addition to the features described above, the GODM system allows resource managers and other users to manage their own projects (individual datasets) with high security, convenience, and dependability. Experts can add photos, textual information, and Internet "Links" to the 'species profiles' and other areas of the web site. Bulletin boards provide information local to an area or a specific invasive organism, while an "early warning" system will provide emails to users as

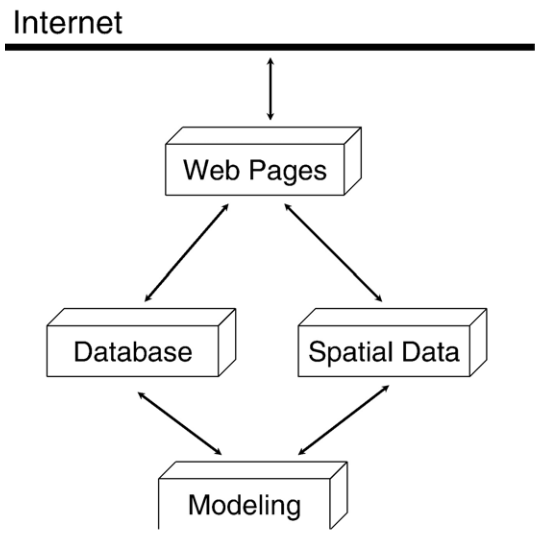

Fig. 1 – The four major subsystems of the Global Organism Detection and Monitoring system (GODM). Web Pages provide dynamic access to the database and available spatial data. Modeling provides predictive maps.

new species invade. Dynamic maps allow users to examine data at the finest resolution, which may only include a single *Elaeagnus angustifolia* (Russian olive) tree along a river, to large pastures infested with *Euphorbia esula* (leafy spurge), to the global distribution of a plant genus such as *Tamarix* (salt cedar). The maps display appropriate background images including standard United States Geological Survey topographic maps and political boundaries to help users locate data. The map application integrates maps available from other Internet sites through the Web Mapping Service (WMS) protocol.

There are four major sub-systems of the Global Organism Detection and Monitoring system (GODM; Fig. 1). The web pages subsystem provides access to the information in the database and access to spatial data. The database maintains biological and other information, while the spatial data subsystem manages the large vector and raster datasets required. The modeling subsystem will provide predictive maps and will accommodate an expanding set of modeling methods. The system is connected to the Internet allowing users from virtually anywhere in the world access using standard Internet browsers such as Internet Explorer and Netscape Navigator (www.NIISS.org).

2. Database design

The database was designed to provide a high level of flexibility while maintaining performance. It is implemented as a relational database in Microsoft SQL Server 2000. The database design is organized into the following relational components:

- Core Survey Data: Data from field surveys
- Survey Addition: Structure of files from field surveys
- Data Addition: Metadata on an addition of field data to the database
- Data Tables: Online spreadsheets extracted from the field data
- Spatial Data: Raster and vector data in various projections
- Map Configuration: Settings to draw maps
- Taxonomic Information: Taxonomic classifications and profiles
- User Information: Logins, expertise, and project affiliations.

The core survey database tables for recording invasive species survey data are organized into three categories; “Where, Who & When, and What” (Fig. 2). “Where” includes the area surveyed and its associated spatial data. Spatial data could be a point, a series of line segments for a river, or a series

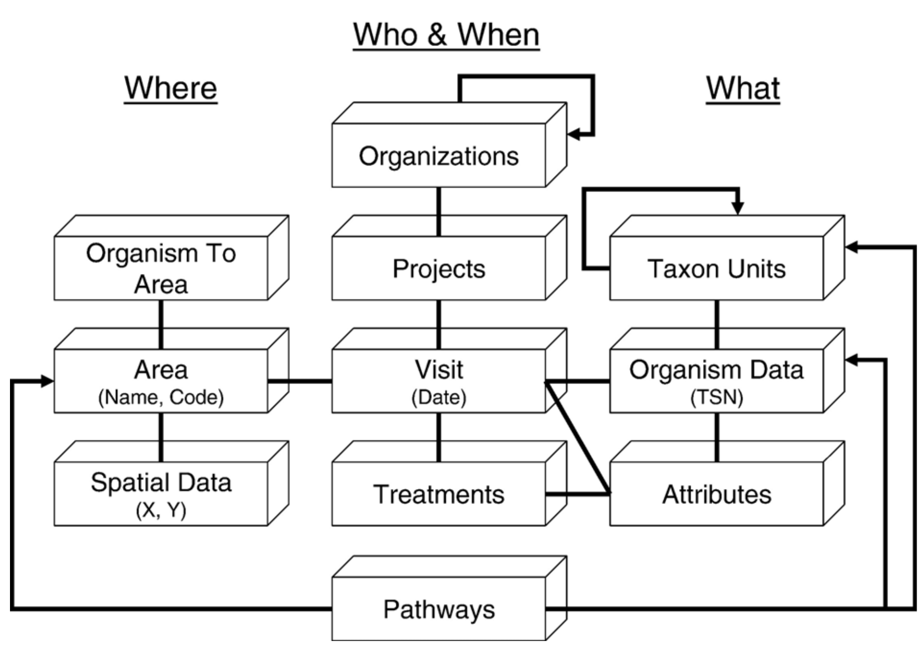

Fig. 2 – Database tables for organism survey data.

of polygons (delineated patches of a species or an area the organism was found in) for a physical region. “Who & When” includes organizations with projects, visits, treatments, and pathways. Organizations include individuals, government agencies, and non-government organizations. A visit represents a survey on a specific date, at a specific location, and who completed the survey. Treatments track invasive species control efforts to determine what control efforts work best in certain environments for specified costs. “What” includes taxonomic information (taxon units), data on the organisms found (organism data) and attributes of the organisms (attributes). Organism data represents the occurrence of an invasive species or associated species at a visit. Organisms are tracked with Taxonomic Serial Numbers (TSN) from the Integrated Taxonomic Information System (ITIS). Attributes include the height of invasive plants; the abundance of animals or diseases; the life stage for insects; or the specifics of a treatment (e.g., the concentration of a particular herbicide or pesticide). Visits can be linked to abiotic attributes (e.g., elevation, soil type, host species). In the future, pathways (seed dispersal by birds, ballast water, and interstate commerce) will allow recording of the mechanism that allowed an organism to arrive at a new habitat (Fig. 2).

To cover the wide variety of ways that organizations collect information, GODM allows users to describe the configuration of their data files for addition to the database. This information is saved in the survey addition section so, if a user has several files that share the same structure, they do not have to re-describe the configuration. Users can also download field collection tools, called EcoNab, to quickly upload files without having to define specific file configurations. Users can store their associated metadata and citations in the data additions section.

Data providers or users may query the database with a flexible query engine. These data are stored in data tables that can be used for statistical analysis, modeling, and downloading.

GODM allows users to customize maps with data on different organisms and queries based on a variety of environmental attributes. They can add informational layers from the spatial data subsystem or from other web services. The spatial data subsystem contains vector data (e.g., points, polygons) and the locations of files containing raster data (e.g., remote sensing outputs, maps). Customized maps are then saved in the map configuration section.

Users can search invasive species information by scientific name and common name. Natural Resource Conservation Service (NRCS) codes can also be used to search for plants. Taxonomic information includes a relational hierarchy across taxa from the ITIS.

To add data to the database, or to access advanced features, users must be registered with GODM. During or after registration, users can request a higher level of security to become project managers or add information as an expert. As a project manager, they can authorize other users to add and edit data for their project. Certain data can be classified as sensitive and can only be viewed by the data provider/project manager or by users that obtain a special clearance from the project manager (Jarnevich et al., in press). Information on users and their capabilities is maintained in the user information section.

The remainder of the database comprises over 100 tables. These contain more detailed data on map projections used, ancillary environmental data, source contact information, level of security, taxa identification certainty, data location, sensitivity, and many fields for analysis and modeling. A full representation of the database schema along with a detailed database dictionary is available at www.niiss.org.

3. Web interface

The Global Organism Detection and Monitoring system is specifically designed to provide “living maps” of harmful invasive species. For example, policy makers requested an updated map of known locations for the invasive riparian plant species, *Tamarix*, within the United States. Data were obtained from over 100 projects across the United States and uploaded into GODM creating the most complete map of *Tamarix* locations to date (Fig. 3). Details for each of the projects are available at www.niiss.org.

Users can zoom in to USGS Quadrangle maps or zoom out to the world, print maps, and download data. The system can be integrated with existing web sites with a custom “skin.”

GODM is being developed with Microsoft Corporation’s Server 2003 operating system and Microsoft Corporation’s Internet Information Server as a web server. Wherever possible the programming was done in PHP. PHP provides high-level object-oriented programming, has very high-performance for a scripting language, and allows portability to other computer hardware and software. Custom Geographic Information System (GIS) libraries were created where required for processing large datasets, accessing the operating system, or to use existing libraries.

We designed the software using object-oriented methods to speed the addition of features and minimize maintenance of the system (Kamath et al., 1993). Quality has been assured using industry standard software testing methodology (Jacobson et al., 1998).

4. Spatial data

To allow users to upload a file of data with locations of invasive species and immediately see a map of their data combined with other user’s data, a custom GIS Internet solution was required. The GIS solution provides an unprecedented level of flexibility in database access, allowing users to display maps of invasive species distributions or abundances based on various criteria including taxonomic classification (i.e., phylum or division, order, class, family, genus, species, subspecies, and variety), a specific project, a range of dates, and a range of attributes (percent cover, age, height, sex, weight). This is a significant paradigm shift from “map servers” to true Internet-based GIS solutions.

Vector data such as points, multi-segment lines, and polygons are stored in the database with lines and polygons compressed into a binary large object (BLOB). Both remotely sensed and GIS-based raster datasets are compressed using Enhanced Compressed Wavelet (ECW, 2006) files using libraries from ERMapper. ECW allows high-performance viewing of

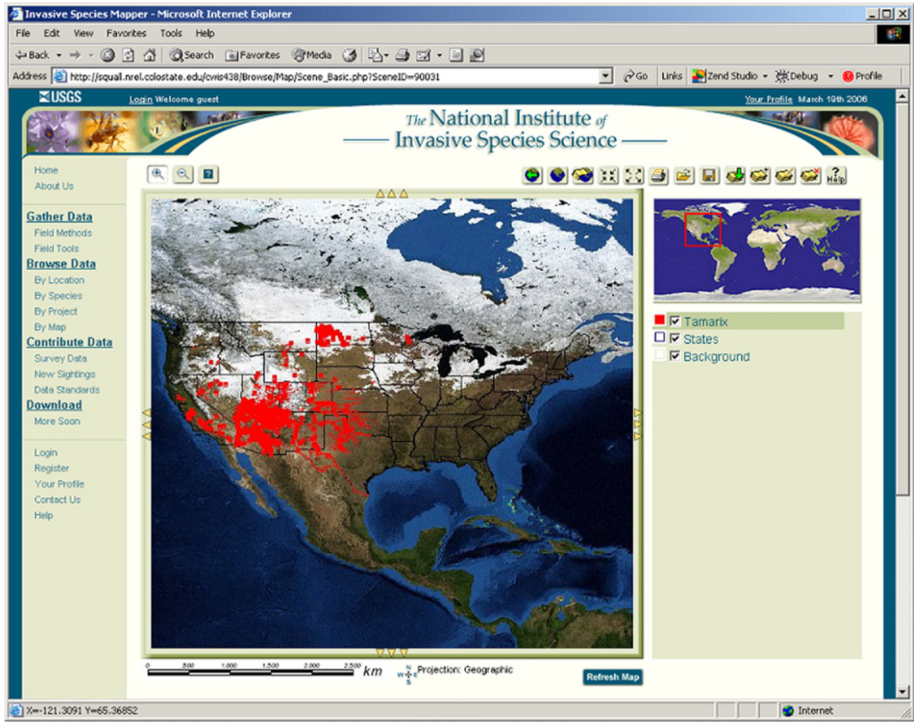

Fig. 3 – The map viewer page showing the actual current data on the distribution of *Tamarix* in the United States from over 100 independent databases.

raster data at virtually any scale and extent. For analysis, the original data is available in uncompressed Tagged Image File Format (TIFF).

To allow the user to view the entire earth and zoom to small areas, the GIS solution needed to provide displays in various projections with minimal delay. The points, multi-segment lines and polygonal data that represent various types of locations are stored in the Geographic projection and in the Universal Transverse Mercator (UTM) projections zones that the data overlap with. Raster data layers are also provided in various projections as needed.

The custom GIS solution was written in C++ and contains a large number of open-source components. Users can upload and download data in a variety of projections and datums. Proj4 (Evenden, 1990) is used to project this data into the Geographic and UTM projections in the World Geodetic System 1984 (WGS84) datum. The Geospatial Data Abstraction Library (GDAL) provides the ability to read and write projection files. Various open-source file translation libraries are used to read and write data files.

Raster datasets such as the United States Geological Survey's 1:24,000 scale topographic maps can contain over 10,000 files in a single UTM zone. To maintain performance requirements, grids were overlaid on these file sets. The database stores the rows and cells of spatial data from the grids, allowing database indexing to restore search times (Longley et al., 2001).

5. Integration of modeling

The GODM system provides an extensible architecture to support many existing modeling methods (e.g., Morisette et al., 2006) and the ability to add new methods in the future. GODM is complimentary to systems such as the NASA Invasive Species Forecasting System (Schnase et al., 2002).

The open source statistical package, "R", has been integrated into GODM to provide a wide range of statistical calculations which can be used along with raster layers of the predictor variables to generate predictive surfaces. Predicted surfaces can then be added to a map with observed locations to evaluate results spatially. These surfaces are the products resource mangers desire to prioritize and direct field efforts to control invasive species. In the future the open source Geospatial Statistical Library (GSLIB) will provide spatial statistics such as semi-variograms and Kriged surfaces. These future changes will allow GODM to bring a far greater array of high-performance predictive spatial modeling tools to our variety of user communities.

6. Hardware architecture

The hardware for the GODM system has been designed to meet the high-performance requirements of performing visualizations and analysis on large geospatial data sets (Fig. 4). Requests are entered through the Internet and are routed to a Load Balancer on the main system or a mirrored system. The Load Balancer sends requests to an idle Web Server to insure users have a quick response. Long processing jobs are submitted to the database and are picked up by Compute Servers. The GIS Servers provide high-speed access to large raster datasets.

To insure performance and reliability the entire system will be mirrored at multiple locations. The web servers will

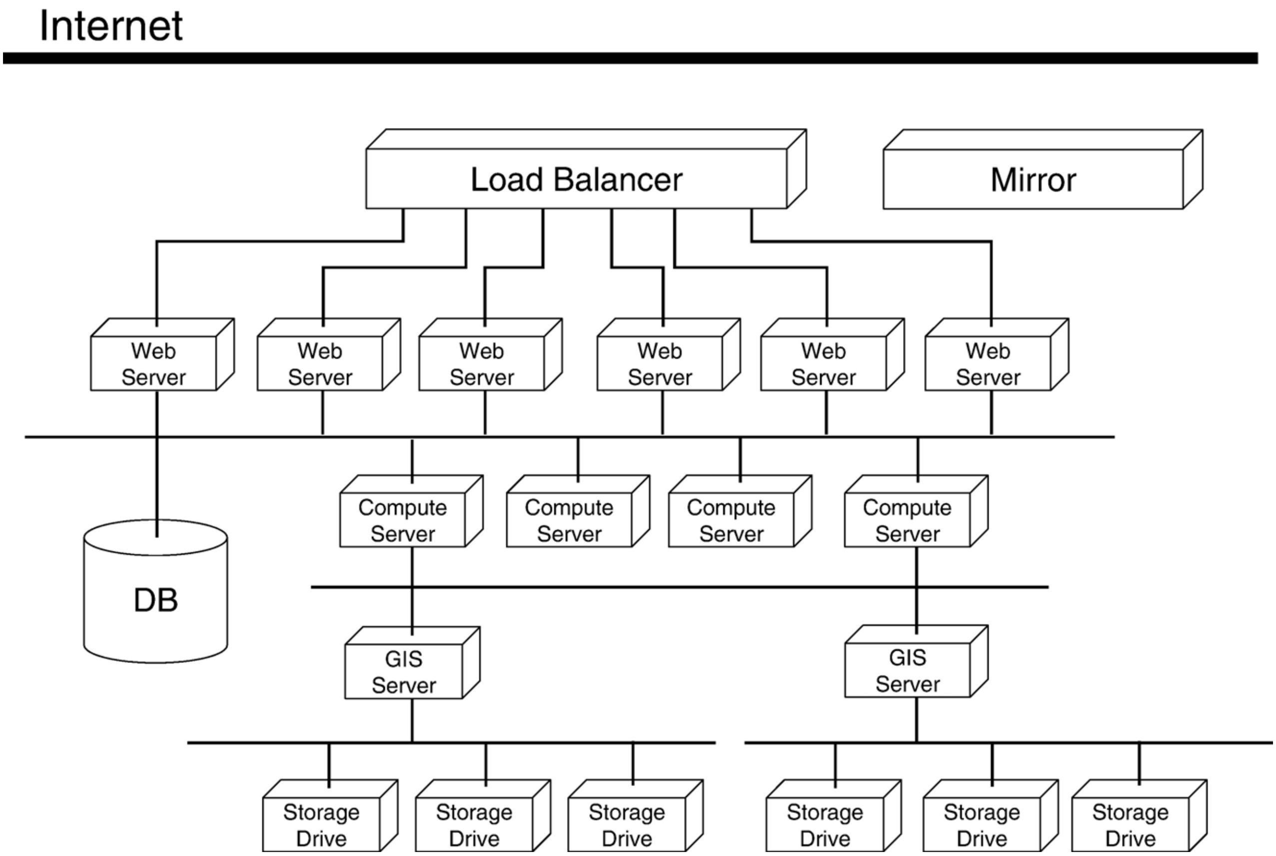

Fig. 4 – Hardware components for the Global Organism Detection and Monitoring system.

be housed at the United States Geological Survey's Fort Collins Science Center and the Natural and Environmental Sciences Building at Colorado State University.

7. General application across the globe

The GODM system represents a significant shift in the way invasive species information is managed and shared. GODM will allow academics, resource mangers and the public to share data and participate in group analysis and decision-making about the way we manage biological resources, and especially harmful invasive species. Sharing data on the web will improve our capabilities to synthesize information on harmful, non-native species (Stohlgren and Schnase, 2006) and improve local management actions (Barnett et al., in press). The technology is applicable to other areas of natural resource management including the management of threatened and endangered species, fire management, and for tracking wildlife or human diseases.

Future features will include: additional modeling approaches, new statistical methods, additional GIS layers, additional remote sensing layers for more accurate predictive modeling, and a protocol to connect GODM to other databases that contain biological data. Over the next two years, additional modeling strategies for controlling invasive species also will be added. At the time of publication, the system includes data from 69 projects, including 37,495 field surveys, with over 130,000 organisms identified from 1562 different taxa.

Acknowledgements

This research was funded by a NASA grant (NRA-03-OES-03); USGS Invasive Species Program; National Biological Information Infrastructure; and Colorado Agriculture Experiment Station. We received logistical support from USGS Fort Collins Science Center and the Natural Resource Ecology Laboratory at Colorado State University. Hundreds of government and non-government users have contributed data. Over 20 students and research assistants have contributed to the testing of the system. To all we are grateful.

REFERENCES

Barnett, D.T., Stohlgren, T.J., Jarnevich, C.S., Ericson, J.A., Davern, T., Simonson, S., in press. The art and science of weed mapping. Environmental Monitoring and Assessment. [Feb 6, Electronic publication ahead of print].

Bowker, G.C., 2000. Biodiversity datadiversity. Social Studies of Science 30, 643–683.

Crall, A.W., Meyerson, L.A., Stohlgren, T.J., Jarnevich, C.S., Newman, G.J., Graham, J.J., 2006. Show me the numbers: what data currently exists for non-native species in the U.S.? Frontiers in Ecology and the Environment, in review.

DiGIR. 2007. digir.sourceforge.net. Distributed Generic Information Retrieval.
ECW. 2006. Earth Resources Mapping. www.ermapper.com.
Evenden, G.I., 1990. Cartographic Projection Procedures for the UNIX Environment — A User's Manual, Open-File Report 90-284. United States Department of the Interior Geological Survey, Woods Hole, MA.
Jacobson, I., Booch, G., Rumbaugh, J., 1998. The Unified Software Development Process. Addison Wesley Longman Inc., Reading, Massachusetts.
Jarnevich, C.S. Graham, J., Newman, G., Crall, A., Stohlgren, T., in press. Balancing data sharing requirements for analyses with data sensitivity. Biological Invasions.
Kamath, Y.H., Smilan, R.E., Smith, J.G., 1993. Reaping benefits with object-oriented technology. AT&T Technology 72, 14–24.
Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 2001. Geographic Information System and Science. John Wiley & Sons, Ltd., Chichester, England.
Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M., Bazzaz, F.A., 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10, 689–710.
Morisette, J.T., Jarnevich, C.S., Ullah, A., Cai, W.J., Pedelty, J.A., Gentle, J.E., Stohlgren, T.J., Schnase, J.L., 2006. A tamarisk habitat suitability map for the continental United States. Frontiers in Ecology and the Environment 4, 11–17.
Pimentel, D., Zuniga, R., Morrison, D., 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52, 273–288.
Schnase, J., Stohlgren, T.J., Smith, J.A., 2002. The national invasive species forecasting system: a strategic NASA/USGS partnership to manage biological invasions. Earth Observation Magazine 11, 46–49.
Schnase, J.L., Cushing, J., Frame, M., Forndorf, A., Landis, E., Maier, D., Silberschatz, A., 2003. Information technology challenges of biodiversity and ecosystems informatics. Information Systems 28, 339–345.
Stohlgren, T.J., Schnase, J.L., 2006. Risk analysis for biological hazards: what we need to know about invasive species. Risk Analysis 26, 163–173.
Wilcove, D.S., Rothstein, D., Dubow, J., Phillips, A., Losos, E., 1998. Quantifying threats to imperiled species in the United States. Bioscience 48, 607–615.