

Running a network on a shoestring: the Global Invasive Species Information Network

Catherine S. Jarnevich^{1*}, Annie Simpson², James J. Graham³, Gregory J. Newman⁴ and Charles T. Barger⁵

¹U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Ave Bldg C, Fort Collins, Colorado 80526, USA

²U.S. Geological Survey, Eco-Science Synthesis Branch, Core Science Analytics Synthesis & Libraries Program, 12201 Sunrise Valley Drive, Reston, Virginia 20192, USA

³Environmental Science & Management, Humboldt State University, 1 Harpst St., Arcata, California 95521-8299, USA

⁴Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, Colorado 80523-1499, USA

⁵Center for Invasive Species & Ecosystem Health, University of Georgia, 2360 Rainwater Road, Tifton, Georgia 31793, USA

E-mail: jarnevichc@usgs.gov (CSJ), asimpson@usgs.gov (AS), james.graham@humboldt.edu (JJG), Gregory.Newman@ColoState.Edu (GJN), cbarger@uga.edu (CTB)

*Corresponding author

Received: 17 November 2014 / Accepted: 30 January 2015 / Published online: 26 February 2015

Handling editor: Stelios Katsanevakis

Abstract

The Global Invasive Species Information Network (GISIN) was conceptualized in 2004 to aggregate and disseminate invasive species data in a standardized way. A decade later the GISIN community has implemented a data portal and three of six GISIN data aggregation models in the GISIN data exchange Protocol, including invasive species status information, resource URLs, and occurrence data. The portal is based on a protocol developed by representatives from 15 countries and 27 organizations of the global invasive species information management community. The GISIN has 19 data providers sharing 34,343 species status records, 1,693,073 occurrences, and 15,601 resource URLs. While the GISIN's goal is to be global, much of its data and funding are provided by the United States. Several initiatives use the GISIN as their information backbone, such as the Great Lakes Early Detection Network (GLEDN) and the North American Invasive Species Network (NAISN). Here we share several success stories and organizational challenges that remain.

Key words: information network, invasive species, data aggregator, data standards

Introduction

More than a decade ago researchers realized the need for a global information system for invasive species (Ricciardi et al. 2000). The system would provide necessary information to assist monitoring, risk assessment, and control efforts. In response to this need and with US State Department funding, the US Geological Survey held a meeting in 2004 with participants from 26 nations forming a Global Invasive Species Information Network (GISIN). Outcomes of the meeting included the Baltimore Declaration (GISIN 2004) which outlined the foundation of the GISIN community and cyberinfrastructure, the compilation of The GISIN List (<http://www.gisin.org/GISINList.htm>) of online invasive species information systems that

could become data providers, the white paper *Databasing Invasions* (Sellers 2004), and increased awareness among participants of the need for effective data sharing to support decision makers (Sellers et al. 2004).

A decade later, in spite of financial and organizational obstacles, <http://www.gisin.org> continues to provide a global platform for invasive species information sharing via the Internet. The GISIN is an informal network of invasive species experts, researchers, information managers, and computer scientists sharing their knowledge and experience to improve access to information needed for prevention, management, control, and research on invasive species. This network collaboratively developed a protocol for exchanging invasive species data. The GISIN data exchange Protocol (hereafter referred to as the GISIN Protocol)

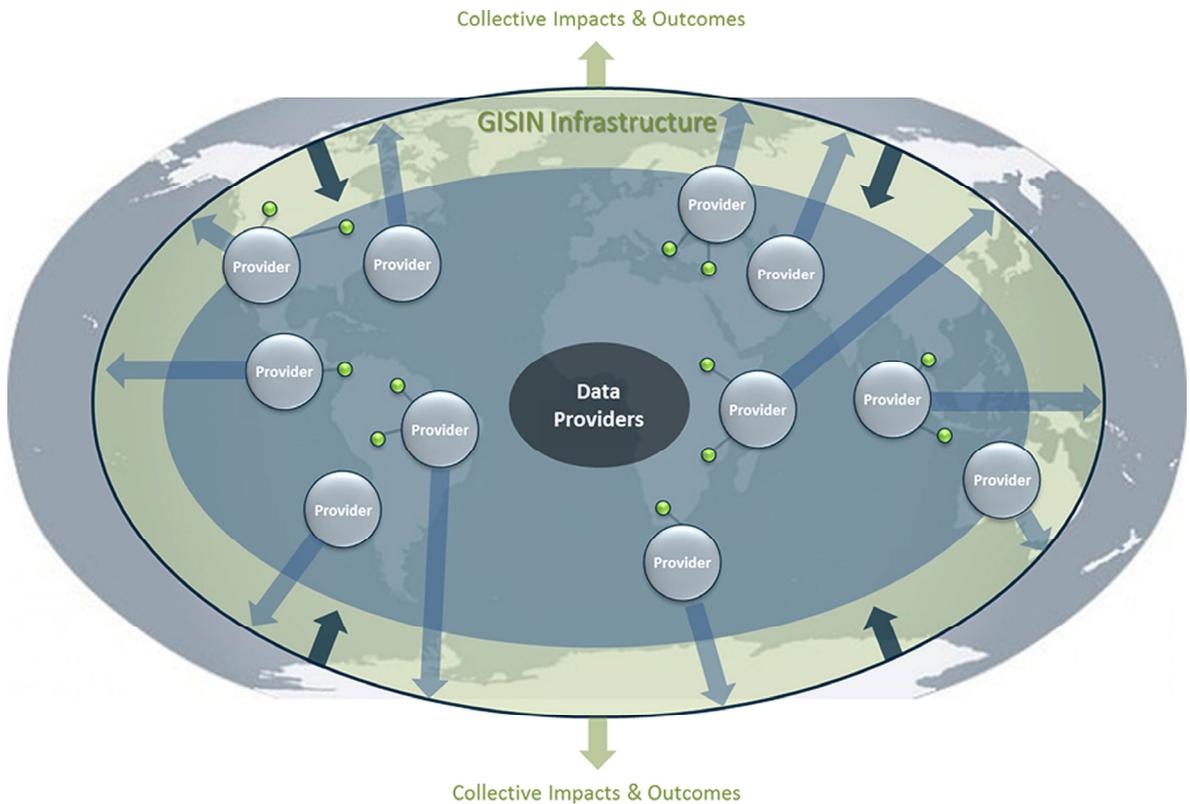


Figure 1. Overall architecture of the GISIN Infrastructure. Various data providers share data to the GISIN portal (lighter blue wide arrows sharing data out to the GISIN network). All data providers and stakeholders can in turn query data from the GISIN portal (darker blue wide arrows sharing data back in towards central stakeholders). Green dots represent data providers who may be themselves networked and sharing information to a more regional data provider or aggregator.

allows these data to be easily aggregated and disseminated, rather than requiring that individuals visit multiple sources and manually aggregate data themselves (Figure 1). The GISIN data are combined from multiple data providers while retaining ownership, attribution, and clear ties to original data providers. Currently, data providers add data to the GISIN cyberinfrastructure through file uploads or web services, and anyone may obtain aggregated data on invasive species locations, status in a particular area, and profile or fact sheet information, using either file downloads or web service queries. Several other data aggregation efforts exist, but these are generally for a smaller geographic area, are not focused only on invasive species, and/or focus on a subset of the data types the GISIN provides (such as species occurrence information; e.g., see review by Crall et al. 2006).

In 2008 the GISIN principals presented a vision for a cyberinfrastructure for invasive species data (Graham et al. 2008): providing data

on locations and characteristics of invasive or potentially invasive species; watch lists of potential invaders for specific geographic areas; early detection alerts; models of current and predicted ranges; and data regarding species best management practices. Here we give a brief overview of the current state of the GISIN and its partners and an assessment of how the GISIN is implementing this cyberinfrastructure.

The GISIN has partnered with the North American Invasive Species Network (NAISN, <http://www.naisn.org>) and the Great Lakes Early Detection Network (GLEDN, <http://www.gledn.org>) to maintain and expand its infrastructure based at Colorado State University in Fort Collins. The GISIN also relies on limited funding and in-kind support from the US Geological Survey (USGS) and in-kind support from Colorado State's Natural Resource Ecology Laboratory, the University of Georgia's Center for Invasive Species and Ecosystem Health, and Humboldt State University.

Over 200 registered invasive species databases were included in The GISIN List in 2004. These websites contain a range of disparate information on invasive species including species profiles, images, locations, distribution information, and species status in different locations throughout the world. Thus, searching for information on invasive species requires examining a large number of potential resources. Although there is a large amount of data available, it is difficult to find and make interoperable and comparable across resources. More recently, in the US alone over 300 databases containing invasive species information have been identified, several of which are not on-line (Crall et al. 2006).

GISIN Architecture and Technical Challenges

Specification

In 2007 the GISIN team members conducted a needs assessment survey of the invasive species information management community, obtaining 137 responses from 41 countries (Simpson et al. 2007). The results from this survey informed the architectural development of the GISIN. A majority of the survey respondents had limited technical knowledge about how to share their data effectively. This meant that any data sharing system such as the GISIN had to be relatively easy to use with minimal training.

Graham et al. (2011) compared a federated system, where each data provider was searched every time a search request was made, to a cached system, where information was aggregated and cached in a central location for searching. Federated searches performed poorly and were inflexible compared to a cache, leading the authors to conclude that a cached system is necessary to support a system such as the GISIN, with data from providers aggregated into a central location and available via a portal. A federated solution would also potentially exclude data providers with limited technical expertise and infrastructure. Given these findings, the GISIN was developed following the cache model and now serves as the interface connecting many disparate data providers that are often in turn connected to several regional and national data repositories (Figure 1).

GISIN Protocol

For groups to effectively share data, there must be a common data sharing protocol. The GISIN protocol, which is compliant with the Darwin

Core body of standards (<http://rs.tdwg.org/dwc/>), consists of six data models developed by a broad spectrum of members of the invasive species information management community over a period of three years (Convention on Biodiversity 2006; Sellers and Simpson 2008a, b). The meetings held to create the GISIN protocol included representation from 30 organizations and 15 countries (Table 1). In all, three GISIN technical meetings were held to determine the six data models and the concepts within them (Table 2). The full GISIN protocol is available on the GISIN website at http://www.gisin.org/cwis438/websites/GISINDirectory/Tech/Protocol_Home.php

Portal

To facilitate data sharing from as many groups as possible, the GISIN cyberinfrastructure includes several paths of varying complexity for data to be shared through the portal. Those with small data sets are directed to collaborate with existing data providers in their area based on type of data, taxonomic group, and geographic location. For groups with infrequently updated data sets and minimal information technology support, there is a file upload capability. The file upload process allows users to map the fields in their data set to standardized GISIN data concepts. For those with large, frequently-updated datasets with strong information technology support, a web services toolkit is provided (GISIN 2015). Data providers can install the toolkit on their local computers and map the values in their database to the standard values defined in the GISIN protocol. The data providers can then configure the GISIN cyberinfrastructure to upload their data on a one-time or periodic basis.

Data quality assurance and quality control

The quality of data from a diverse group of data providers can vary greatly (Chapman 2005; Hellerstein 2008). Data quality checks include ensuring that required fields are provided and that the values within each data field are valid. The GISIN portal checks for correctly formatted geographic coordinates, other numbers, and dates. The GISIN protocol also makes use of standard controlled vocabularies for country names (two letters codes from the International Organization for Standardization, ISO 3166) and languages (language syntax tags from best current practice, RFC 4646), defines new vocabularies for GISIN-specific concepts such as ‘dateOfIntroductionPrecision’ (values of Unknown, Day, Month, Year),

Table 1. Organizations participating in the creation of the GISIN Protocol.

Organization	Country
ASEANET - The Southeast Asian LOOP of BioNET-INTERNATIONAL	Malaysia
Berlin Botanical Gardens, Germany (for TAPIR)	Germany
BioNET-INTERNATIONAL	Great Britain
Centre for Biological Information Technology	Australia
Centro de Referência em Informação Ambiental (CRIA)	Brazil
Cooperative Research Centre for National Plant Biosecurity	Australia
Delivering Alien Invasive Species Inventories for Europe (DAISIE)	Germany
Directorate of Inspection and Food Quality	Morocco
Directorate of Plant Protection and Technical Inspections	Morocco
Discover Life	United States
ETI BioInformatics	Netherlands
FISHbase	Philippines
Global Biodiversity Information Facility	Denmark
Global Invasive Species Database	New Zealand
Great Lake Indian Fish and Wildlife Commission	United States
Hawaiian Ecosystems at Risk	United States
Institut National Agronomique de Tunisie	Tunisia
IABIN's Invasives Information Network (I3N)	Argentina
IUCN-Invasive Species Specialist Group	New Zealand
NASA, Goddard Space Flight Center	United States
National Biological Information Infrastructure	United States
National Institute of Invasive Species Science	United States
NISbase	United States
Pacific Island Ecosystems at Risk	United States
Secretariat of the Convention on Biological Diversity	Canada
Smithsonian Institution	United States
Universidad Nacional del Sur (INBIAR)	Argentina
University of Georgia	United States
University of Massachusetts	United States
US Geological Survey Nonindigenous Aquatic Species database (NAS)	United States

Table 2. GISIN data exchange protocol data models.

Model Name	Status	Description
SpeciesStatus	Implemented	Represents the status of a species in a given area, including whether the species is native or non-native and whether it causes harm.
SpeciesResourceURL	Implemented	Contains URLs (web links) to resources for an invasive species, including species profile pages, images, and other content.
Occurrence	Implemented	Contains records of a species, date, and location in which the species was observed.
ManagementStatus	Defined	Contains information on how an invasive species is being managed in a given area.
ImpactStatus	Defined	Describes the types of impacts a species is having in an area.
DispersalStatus	Defined	Provides information on where a species came from, how it was introduced, and how it is spreading through an area.

For more detail, see: http://www.gisin.org/cwis438/websites/GISINDirectory/Tech/Protocol_Home.php

and allows the addition of new vocabularies that are country specific. These controlled vocabularies cross-walk various contributed names of values into unique identifiers to facilitate comparability across data from different data providers. These types of checks are not uncommon in data aggregators (e.g., Belbin 2013).

When data are uploaded by data providers, each record is immediately checked and any records with errors are not added to the GISIN

cache. Data providers are given a list of data rows that contain errors, with the errors highlighted. This feedback can provide data quality checks for data providers, who can correct the problems and re-upload the data set to the GISIN cache.

Integrated data may contain records shared by multiple data providers, generating duplicates. For example, the International Biological Information System (IBIS; hosted at Colorado State University) and Early Detection and Distribution Mapping

Table 3. GISIN data providers.

Provider Name	Provider short name	Species Status	Occurrence	Species Resource URL
Center for Aquatic and Invasive Plants	UF-IFAS CAIP			111
Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (National Commission for Knowledge and Use of Biodiversity)	CONABIO		14020	
Delivering Alien Invasive Species Inventories for Europe	DAISIE	7699		37
Early Detection and Distribution Mapping System Data	EDDMapS		908582	
FishBase	FishBase		212301	
Geosystems Research Institute	GRI		16717	
Global Invasive Species Database	GISD	17422		311
Global Register of Invasive Species	GRIS	158		
Great Lakes Indian Fish and Wildlife Commission	GLIFWC	8	8735	
Hawaiian Ecosystems at Risk project	HEAR			20
Invasives Information Network -Argentina	I3N-Argentina		654	
Illinois Database of Aquatic Non-native Species, Loyola University Chicago	IDANS		26061	
International Biological Information System	IBIS		299222	397
Midwest Invasive Species Information Network	MISIN		41643	
Nonindigenous Aquatic Species Program	NAS	8592	155318	14424
Northwoods Cooperative Weed Management Area	Northwoods CWMA		9132	
Taiwan Forestry Research Institute	TFRI	163		
The Calflora Database	Calflora	301		301
Towson University - MidAtlantic - <i>Ailanthus altissima</i> - 2011			688	
Totals:		34343	1693073	15601

System (EDDMapS; hosted at University of Georgia) data providers aggregate data from some of the same sources. Both data providers share these data with the GISIN, but the duplicates are identified and removed from the GISIN cache, using Globally Unique Identifiers (GUIDs) for detecting and eliminating redundancy in the data. GUIDs for data records are available from GUID authorities (including the GISIN) and, when used, can be disambiguated no matter who holds the data. GUIDs also allow the original data provider of the data to be identified. However, despite their utility, GUIDs are not yet commonly included in datasets, and thus are not a required field to share data with the GISIN. In an effort to promote their use, the GISIN website does provide information to assist in generating GUIDs following the standard [Authority]/guid/[InstitutionCode]/[CollectionCode]/[CatalogNumber], where the GISIN may be used as an authority.

Examples of Integration

The GISIN cache – aggregated invasive species data

The GISIN cyberinfrastructure vision included collection of occurrence locations and characteristics of invasive or potentially invasive species. The three implemented data models of the GISIN protocol (Species Status, Species Resource URL, Occurrence) do this, with data in the cache from 19 data providers (Tables 2 and 3).

Currently, the GISIN cache includes seven data providers sharing species status data for 1,847 species. There are four species with status information contributed by five data providers, 19 with status information from 19 data providers, 102 with status information from three data providers, 350 with status information from two data providers, and 1,371 species with status information provided by only one data provider. By aggregating data from different data providers we obtain a more complete picture of a species'

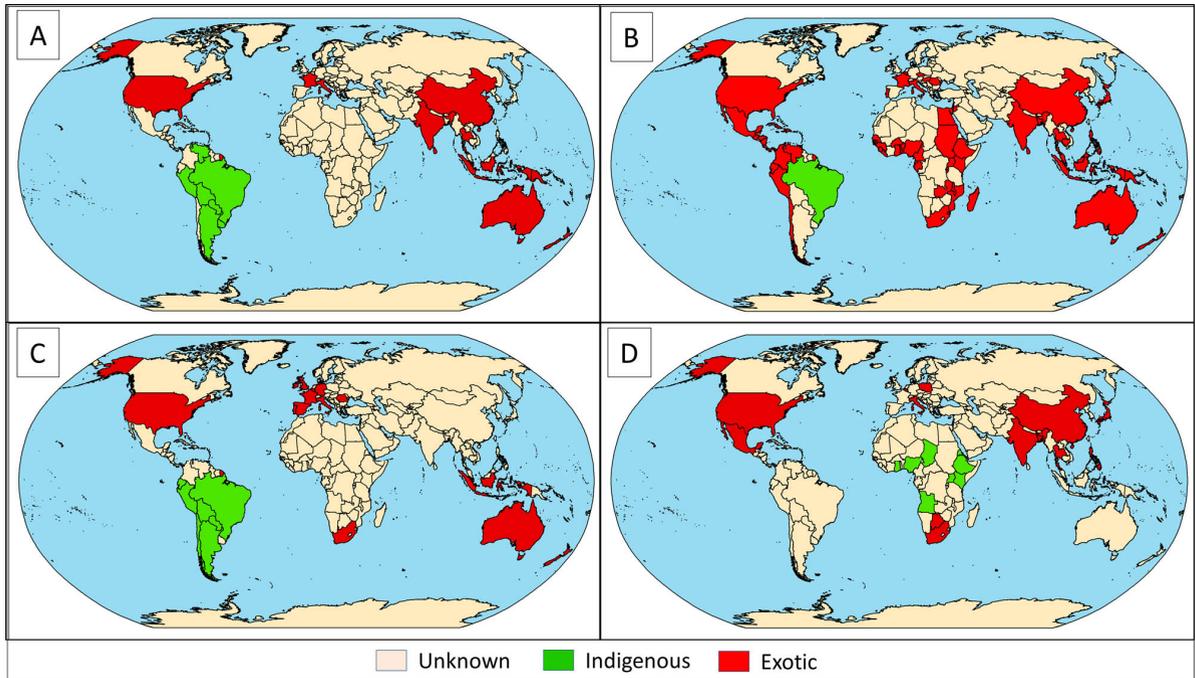


Figure 2. Species status information contributed by five data providers for four different species including A) *Alternanthera philoxeroides*, B) *Eichhornia crassipes*, C) *Myriophyllum aquaticum*, and D) *Oreochromis niloticus*.

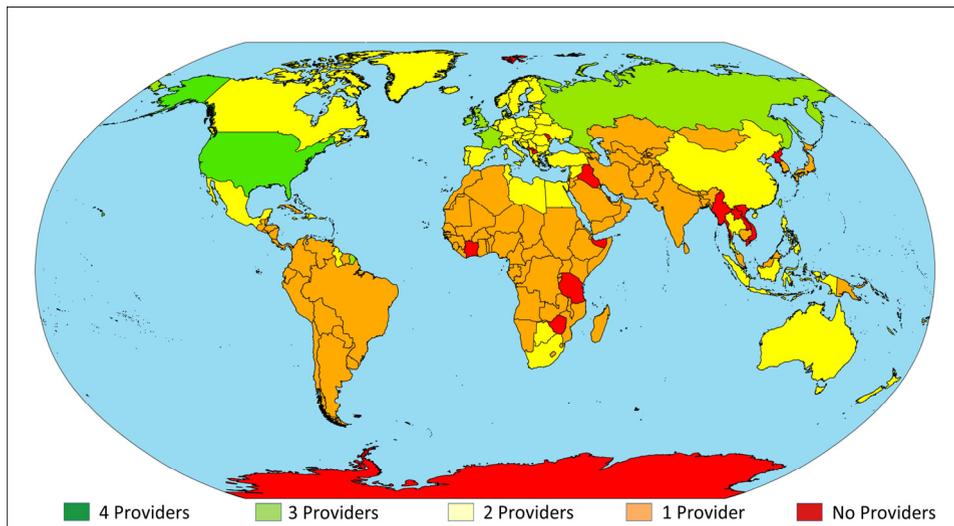


Figure 3. Number of data providers contributing species status data by country.

status globally, as with the four species with data shared by five data providers (Figure 2). There is broad global coverage, with almost all countries having at least some information from at least one data provider (Figure 3). Within the Species Resource URL data model there are seven (other

data providers, with resource URLs for 2,484 species. There are four species with species resource URLs from four data providers, 18 with URLs from three data providers, two with URLs from 115 data providers, and 2,347 with URLs from a single data provider. Six of the seven data

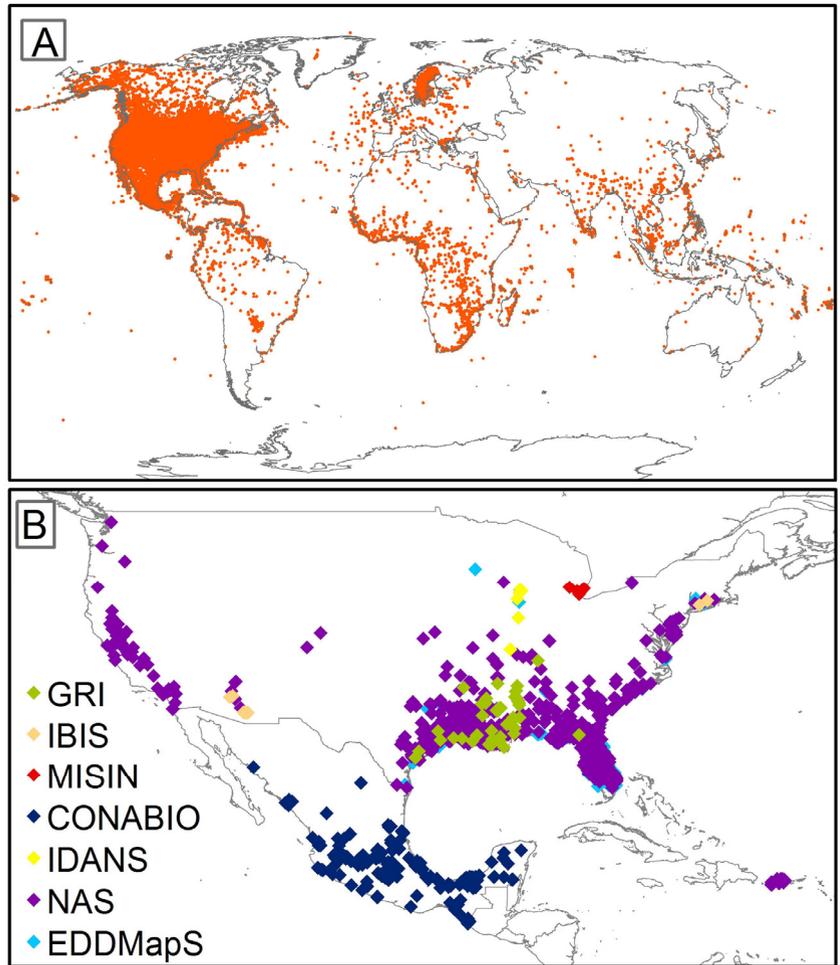


Figure 4. Map showing A) all species occurrence locations shared through the GISIN portal and B) occurrence locations for *Eichhornia crassipes* from eight different data providers (see Table 2 for more information on the data providers).

providers contributed a URL for a species not represented by the other data providers. Thus, by using the GISIN portal people searching for a specific species have a higher chance of obtaining a URL for the species than by going to an individual website.

The Occurrence data model includes 12 data providers contributing almost 1,700,000 records for 4,697 species from six taxonomic Kingdoms. Three species had data from eight data providers, 12 from seven, with continued increasing numbers to 3,432 species with location data shared by a single data provider. These data cover most of the globe, but there is an obvious bias to North America (Figure 4a). By merging data from multiple data providers there is better, more comprehensive representation of a species' distribution than that represented in the data of any single data provider (Figure 4b).

Sustainability Issues

Initial funding for the GISIN was provided by the US Department of State Bureau of Oceans and International Environmental and Scientific Affairs to the USGS. This was followed by travel grants from the Group on Earth Observations, NASA, and the Secretariat of the Convention on Biological Diversity, and funding for software development from the USGS in a cooperative agreement with Colorado State University that ended in 2011. A global data partnership to organize and manage invasive species information also includes a large amount of in-kind support from organizations and individuals with supportive leadership who share common goals.

The survival of an organization that is a virtual network with a voluntary secretariat is tenuous at best. It depends on the passion and dedication of

individuals, often in spite of a lack of funds and other resources. When necessary, such organizations may need to change their emphasis to accommodate the type of funding available. Although the GISIN retains a global scope, its most recent IT development has a North American focus, with funding from the Commission for Environmental Cooperation (CEC), which has a tri-national mandate. The CEC has provided financial support to the GISIN via its partner organization NAISN, briefly described below. In total, GISIN received less than one million USD in grants over a 10 year period, and currently does not receive any direct funding.

GISIN end users

The GISIN data portal is an integral component of many local, regional, and nationally focused web application systems within the United States and elsewhere. In each case, the GISIN portal serves as the link between a network of data providers and the web application. Data is harvested from a network of data providers via the GISIN protocol to support specific, end-user relevant goals. One goal described in the vision of cyberinfrastructure was the ability to send alerts regarding early detection of new invaders.

The GLEDN, a regional invasive species network focusing on the Midwestern United States, automates the delivery of customized email alerts for stakeholders, where stakeholders are notified of a new species occurrence meeting their criteria (either species of interest or geographic area of interest; Crall et al. 2012). The GISIN protocol serves as the backbone of the GLEDN system, providing the mechanism to aggregate data across data providers.

The NAISN is an American 501(c)3 non-profit organization that was formed in 2010 by university and government scientists from across North America. Organizations from Mexico and Canada participate as NAISN members through a Memorandum of Understanding. Currently six regional “hubs” (Colorado State University, University of Florida, University of Georgia, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Texas Invasive Species Institute and the Canada-Ontario Invasive Species Centre) and two thematically focused “nodes” (Montana State University and Algoma University) make up NAISN (NAISN 2015). Because invasive species are not restricted by governmental jurisdictional boundaries, NAISN aims to unify and connect existing regional invasive species efforts into a

single network to improve communication, collaboration, and overall coordination on current invasive species management and prevention efforts. One early decision made by NAISN was to fully endorse, support and promote the use of the GISIN cyberinfrastructure.

EDDMapS, another GISIN partner, was originally developed in 2005 to map invasive plants in the southeastern United States by the University of Georgia’s Center for Invasive Species and Ecosystem Health. Ten years later, EDDMapS has become a platform for reporting and mapping all-taxa invasive species and agricultural pest distribution in the United States and Canada. EDDMapS provides web-based reporting forms, JavaScript based state and county distribution maps, and point distribution maps using both Google Maps application programming interface and ESRI’s ArcGIS for Server. These maps and reporting forms are available through regional portals and embedded on cooperators web sites (e.g. <https://njaes.rutgers.edu/stinkbug/report.asp>). EDDMapS also provides sixteen smartphone applications (<http://apps.bugwood.org>) for Apple iPhone, Apple iPad and Android platforms. The CEC provided funds to incorporate the University of Georgia’s EDDMapS data set of over 900,000 occurrence records into the GISIN cache and provide an Internet mirror of the GISIN cache as part of the NAISN and EDDMapS websites. EDDMapS provides its users with tools for quickly visualizing species distribution(s) through one centralized web portal.

BISON (Biodiversity Information Serving Our Nation, <http://bison.usgs.ornl.gov>), launched by the USGS in 2013, is another GISIN partner that maps species occurrences throughout the US and its Territories, with more than 160 million records and 50 map layers. BISON emphasizes the mobilization of federal and other species occurrence data (including invasive species data sets), which is shared with the GISIN cyberinfrastructure and vice versa.

Successes

A major impediment to the GISIN's success is some data holders' unwillingness to share their data. The GISIN has provided some incentives for data sharing. All data provided to the GISIN cyberinfrastructure are then available back to everyone, so data holders can increase their data by obtaining the subset of the entire aggregated data set of interest to them. Also, the GISIN cyberinfrastructure provides many links back to

the individual data providers, which can increase Internet traffic and use of their data products. The GISIN protocol is intentionally simple and fairly general, to better enable aggregation, and thus does not capture all of the information available in contributed data sets. Other organizations could link to the GISIN cyberinfrastructure and develop their own portal. For example, the NAISN is developing a GISIN portal specific to North America.

Measuring the success of the GISIN can be difficult. The GISIN provides the backbone for existing successful websites, but its presence is not always obvious to the end user. The GLEDN, NAISN, BISON, and EDDMapS are examples of websites that have integrated the GISIN into their architecture. The GISIN provides an invaluable service in aggregating and disseminating invasive species data from disparate sources to these sites, while operating seamlessly in the background. According to Google Analytics, direct visitation to the GISIN portal as of January 2015 yields the following overall summary statistics: 159 sessions, 130 users, 621 page views, 3.91 pages/session, 02:23 average session duration and a 45.28% bounce rate. We had 71% new sessions the week of January 5.

Researchers used the GISIN cyberinfrastructure to aggregate weed mapping data from ten sources for the state of Wisconsin (Crall et al. 2013). Some of these data sets were smaller, and were contributed to larger data systems that share data with the GISIN cyberinfrastructure. These data were then used to generate habitat suitability models that were used by management agencies to identify locations for more targeted invasive plant surveys, resulting in more detections of unknown populations than 'business as usual' sampling. The effort to aggregate the data used to generate these models was reduced by using the GISIN cyberinfrastructure.

Future directions and conclusions

Future directions for the GISIN include:

- 1) Implementation of the three defined but unimplemented data models
- 2) Adding quality assurance and control checks for scientific names of organisms against available standards (e.g., the Integrated Taxonomic Information System (ITIS))
- 3) Recruiting more data providers – especially reengaging the global community
- 4) Scoping new financial partnerships and strengthening old ones, for stronger sustainability.

The GISIN currently suffers from similar data gaps as other databases and aggregators, with certain taxonomic groups having better coverage (e.g., aquatic species) and certain geographic areas such as the United States being well represented while other areas such as Africa have a paucity of data (Crall et al. 2006; Yesson et al. 2007). As additional GISIN data models are implemented and new data providers join, other possibilities for use of the GISIN protocol and network become achievable such as modeling success of management strategies and analyzing the most critical introduction paths. These possibilities advance science and adaptive management well beyond the mere traditional applications of points on a map; the GISIN protocol facilitates the standardization and subsequent inclusion of more comprehensive, important information about the character and status of invasive species (rather than just their presence on the landscape) into research applications. The vision for an invasive species cyberinfrastructure outlined in 2008 had five main components, most of which the GISIN has at least facilitated, and full implementation of the GISIN data models would further fulfill this vision.

At least two other global invasive species organizations, the Global Invasive Species Programme (GISP) and The Nature Conservancy's Invasive Species Program, have been disbanded for lack of funding. Each had full time staff. The GISIN has continued to exist because of its low overhead (no full time staff) and because of volunteer contributions to it by people who believe in the importance of invasive species information management and sharing.

The GISIN organizational model has proven to be a tenacious one. But without the contributions of governments, international organizations, and volunteers, it would never have been created.

Acknowledgements

We would like to thank the following organizations for their funding and in-kind support, much of it ongoing: US Geological Survey, North American Invasive Species Network, Commission for Environmental Cooperation, Convention on Biological Diversity, US State Department, NASA, Group on Earth Observations, Global Biodiversity Information Facility, Colorado State University, University of Georgia, and Humboldt State University. We are also grateful to the reviewers of the manuscript, who provided valuable suggestions for improvement. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government nor by the GISIN.

References

- Belbin LD, Joanne Hirsch, Tim Hobern, Donald Lasalle, John (2013) A specialist's audit of aggregated occurrence records: An 'aggregator's' perspective. *ZooKeys* 305: 67–76, <http://dx.doi.org/10.3897/zookeys.305.5438>
- Chapman AD (2005) Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen, online at http://www.gbif.org/orc/?doc_id=1262
- Convention on Biodiversity (2006) Progress report on the implementation of the Global Invasive Species Information Network (GISIN). CONFERENCE OF THE PARTIES TO THE CONVENTION ON BIOLOGICAL DIVERSITY Eighth meeting, 7 p. <http://www.cbd.int/doc/meetings/cop/cop-08/information/cop-08-inf-35-en.pdf>
- Crall A, Renz M, Panke B, Newman G, Chapin C, Graham J, Barger C (2012) Developing cost-effective early detection networks for regional invasions. *Biological Invasions* 14: 2461–2469, <http://dx.doi.org/10.1007/s10530-012-0256-3>
- Crall AW, Meyerson LA, Stohlgren TJ, Jarnevich CS, Newman GJ, Graham J (2006) Show me the numbers: what data currently exist for non-native species in the USA? *Frontiers in Ecology and the Environment* 4: 414–418, [http://dx.doi.org/10.1890/1540-9295\(2006\)4\[414:SMTNWD\]2.0.CO;2](http://dx.doi.org/10.1890/1540-9295(2006)4[414:SMTNWD]2.0.CO;2)
- Crall AW, Jarnevich CS, Panke B, Young N, Renz M, Morissette J (2013) Using Habitat Suitability Models to Target Invasive Plant Species Surveys. *Ecological Applications* 23: 60–72, <http://dx.doi.org/10.1890/12-0465.1>
- GISIN (2004) The Baltimore Declaration. Technical Workshop on the Implementation of a Global Invasive Species Information Network (GISIN), Baltimore, Maryland, 6-8 April 2004, 3 p. <http://www.gisin.org/WebContent/WS/GISIN/Documents/BaltimoreDeclaration.pdf>
- GISIN (2015) Web services for the Global Invasive Species Information Network. <http://www.gisin.org/DH.php?WC=WS/GISIN/GISINDirectory/WebServicesForConsumers.html> (Accessed 9 January 2015)
- Graham J, Jarnevich CS, Simpson A, Newman GJ, Stohlgren TJ (2011) Federated or cached searchers: Providing expected performance from multiple invasive species databases. *Frontiers of Earth Science* 5: 111–119, <http://dx.doi.org/10.1007/s11707-011-0152-7>
- Graham J, Simpson A, Crall AW, Jarnevich C, Newman G, Stohlgren T (2008) Vision of a cyberinfrastructure for non-native, invasive species management. *BioScience* 58: 263–268, <http://dx.doi.org/10.1641/B580312>
- Hellerstein JM (2008) Quantitative Data Cleaning for Large Databases. White Paper, United Nations Economic Commission for Europe. <http://db.cs.berkeley.edu/jmh/papers/cleaning-unece.pdf>
- NAISN (2015) North America Invasive Species Network. <http://www.naisn.org/join.html> (Accessed 9 January 2015)
- Ricciardi A, Steiner WWM, Mack RN, Simberloff D (2000) Toward a global information system for invasive species. *Bioscience* 50: 239–244, [http://dx.doi.org/10.1641/0006-3568\(2000\)050\[0239:TAGISF\]2.3.CO;2](http://dx.doi.org/10.1641/0006-3568(2000)050[0239:TAGISF]2.3.CO;2)
- Sellers E (2004) Databasing Invasions: A Review in the Context of the Global Invasive Species Information Network (GISIN). Information International Associates Inc, 34 p. Online at: [http://www.gisin.org/WebContent/WS/GISIN/Documents/Databasing_Invasions_A_Review_ESellers\(IIa\).pdf](http://www.gisin.org/WebContent/WS/GISIN/Documents/Databasing_Invasions_A_Review_ESellers(IIa).pdf)
- Sellers E, Simpson A, Fisher JP, Curd-Hetrick S, Editors (2004) Experts Meeting on Implementation of a Global Invasive Species Information Network (GISIN), Proceedings of a Workshop. 6-8 April, 2004. Baltimore, Maryland, USA. Online at: http://www.gisin.org/WebContent/WS/GISIN/Documents/ProceedingsPDF/GISINProc2004_Intro.pdf
- Sellers E, Simpson A (2008a) Summary report of the second Global Invasive Species Information Network (GISIN) data providers' workshop. 08-12 August 2008 in Elmira, New York, USA. Global Invasive Species Information Network, 6 p. Online at: <http://www.gisin.org/WebContent/WS/GISIN/Documents/GISIN2workshopSummaryAugust08.pdf>
- Sellers E, Simpson A (2008b) Summary report of the first Global Invasive Species Information Network (GISIN) data providers' workshop. 02-05 June 2008 in Athens, Georgia, USA. Global Invasive Species Information Network, 7 p. Online at: <http://www.gisin.org/WebContent/WS/GISIN/Documents/GISINworkshopSummaryJune08.pdf>
- Simpson A, Graham J, Browne M, Saarenmaa H, Sellers E (2007) Results of a Needs Assessment Survey of the Global Invasive Species Information Network (GISIN). In: Proceedings of TDWG. Biodiversity Information Standards (TDWG) Annual Conference 2007, Bratislava, Slovakia. <http://www.tdwg.org/proceedings/article/view/211>
- Yesson C, Brewer PW, Sutton T, Caithness N, Pahwa JS, Burgess M, Gray WA, White RJ, Jones AC, Bisby FA, Culham A. (2007) How Global Is the Global Biodiversity Information Facility? *PLoS ONE* 2: e1124, <http://dx.doi.org/10.1371/journal.pone.0001124>