The U.S. Geological Survey’s Nonindigenous Aquatic Species Database: over thirty years of tracking introduced aquatic species in the United States (and counting)

Pam Fuller* and Matthew E. Neilson

1 U.S. Geological Survey, Southeast Ecological Science Center, 7920 NW 71st St., Gainesville, FL 32653, USA
2 Cherokee Nation Technology Solutions, USGS Southeast Ecological Science Center, 7920 NW 71st St., Gainesville, FL 32653, USA

E-mail: pfuller@usgs.gov (PF), mneilson@usgs.gov (MEN)

* Corresponding author

Received: 22 October 2014 / Accepted: 30 January 2015 / Published online: 23 February 2015

Handling editor: Stelios Katsanevakis

Abstract

The U.S. Geological Survey’s Nonindigenous Aquatic Species (NAS) Database has tracked introductions of freshwater aquatic organisms in the United States for the past four decades. A website provides access to occurrence reports, distribution maps, and fact sheets for more than 1,000 species. The site also includes an on-line reporting system and an alert system for new occurrences. We provide an historical overview of the database, a description of its current capabilities and functionality, and a basic characterization of the data contained within the database.

Key words: aquatic species, United States, database, trends, spatial, invasive, freshwater

Introduction and history

The U.S. Geological Survey (USGS) Nonindigenous Aquatic Species (NAS) Program functions as a repository and clearinghouse for occurrence information on nonindigenous aquatic species from across the United States. The USGS NAS program monitors, records, and analyzes sightings of nonindigenous aquatic species (defined in the program as a taxon not historically native to a watershed; see Richardson et al. 2011 for a review of terminology in invasion biology) throughout the United States. The program also produces email alerts, maps, summary graphs, publications, and other information products to support natural resource managers. Although there are many regional or taxon-specific databases, the narrow scope of these targeted databases can limit knowledge on large-scale patterns and processes influencing nonindigenous species distributions through data fragmentation. The goal of NAS is to provide national perspective on species distributions and introductions and to allow data analysis at a national scale.

In 1978, the U.S. Fish and Wildlife Service’s National Fisheries Research Center in Gainesville, Florida, began monitoring the status and distribution of nonindigenous fish species in U.S. waters. Originally, this information was maintained as paper files, and it was not until 1989 that these records were digitized. Williams and Jennings (1991) gave a description of the early database; Nico and Fuller (1999), Fuller (2001), and Fuller et al. (2013) provided updates. The original goals of the database were to be an information exchange base for monitoring distribution, rate of dispersal, and potential range expansion of established populations (Williams and Jennings 1991). The initiative to maintain scientific information on nationwide occurrences of nonindigenous aquatic species began with the Aquatic Nuisance Species Task Force, a group created by Congress in 1990 with the passage of Nonindigenous Aquatic Species Prevention and Control Act (U.S. Code of Federal Regulations 1990). Since then, the NAS program has maintained the database as a clearinghouse of information for confirmed sightings of nonindigenous aquatic species throughout the United States.
United States to address the need for this type of information by natural resource managers. Through a series of governmental reorganizations (outlined in Fuller et al. 1999), the program is now managed at the USGS Southeast Ecological Science Center in Gainesville, Florida.

**Scope**

The geographic and taxonomic focus of the USGS NAS program has changed throughout the years, with the current primary focus on freshwater animals (fishes, mollusks, crustaceans, amphibians, reptiles, coelenterates, and bryozoans) throughout the United States, the Commonwealth of Puerto Rico, and the territories of Virgin Islands and Guam. Additionally, we also have separate projects that track the Asian Tiger Shrimp (*Penaeus monodon* Fabricius, 1798), and all nonindigenous marine fishes in the U.S. [including the lionfishes *Pterois miles* (Bennett, 1828) and *P. volitans* (Linnaeus, 1758)]. Aquatic plants are not covered as of publication date due to staffing constraints (aquatic plants will be reinstated during 2015). Marine and estuarine species were previously tracked; these data are currently covered by the Smithsonian Environmental Research Center’s Marine Invasions Lab (Fofonoff et al. 2003). We have also designed and maintain, in partnership with the National Oceanic and Atmospheric Administration (NOAA), the Great Lakes Aquatic Nuisance Species Information System (GLANSIS 2014): a region-specific view of the NAS database focused on species of concern within the North American Laurentian Great Lakes region. Taxonomic coverage of the NAS program includes foreign species and those native to North America that have been transported outside of their natural range. All reports of species found outside their native drainages are included regardless of their establishment status, allowing for information about survivability, vectors, propagule pressure, and the origin of an established population that might initially have been considered an isolated occurrence. The temporal range of the NAS database is from 1800 to the present.

**Information management**

**Data sources**

The NAS database obtains data from many sources, including: primary scientific literature; state, federal, and local monitoring programs; museum accession databases; online databases; news feeds; web sites; professional communications; and online public reporting forms. To harness the wealth of biodiversity and biogeographic observation data contained within distributed online databases, NAS staff have developed tools to interact with, retrieve, and process data through web services APIs, beginning with museum collection data available from the Global Biodiversity Information Facility (GBIF 2014). Georeferenced occurrence data are investigated to determine the historical status (native or nonindigenous) for each species tracked by NAS at each location. Determining the status sometimes requires a significant amount of research. Some of the resources used in this determination include: state fish books, Hocutt and Wiley (1986), and species or drainage-specific papers. Occurrence data identified from introduced locations were added to the NAS database through a recently-developed bulk data upload tool. This data mining of GBIF has resulted in more than 12,000 new records being added to the NAS database.

**Online reporting system**

The public plays a significant role in the invasive species issue by acting as “early detectors” of new invasions (Crall et al. 2010, 2011). The NAS system includes online reporting ([http://nas.er.usgs.gov/SightingReport.aspx](http://nas.er.usgs.gov/SightingReport.aspx)), providing users with means of directly reporting species occurrences. The reporting form includes an integrated map to locate the collection area and accurately report geographic coordinates, and allows for photographs to be uploaded and attached to sighting reports for taxonomic verification. This online tool has been a very effective for reporting new invasions, as a recent analysis of our NAS Alert System (described below) showed that 67% of the alerts generated in the past 5 years have come from personal communication through this volunteer reporting mechanism (Fuller et al. 2013).

**Database design**

The NAS database is a relational database comprising 60 fields across several different database tables and includes taxonomic [e.g., classification, common name, authority following the Integrated Taxonomic Information System (ITIS 2014) or more recent publications], geographic (e.g., locality description, geographic coordinates, watershed), and temporal information along with basic data on each individual sighting (e.g., number of individuals, pathway, establishment
Overview and history of the USGS NAS Database

Figure 1. Relational design of the U.S. Geological Survey’s Nonindigenous Aquatic Species database. Boxed indicate primary data tables used for information storage, retrieval, and presentation. Arrows indicate relationship among tables within the database, with primary identifiers (i.e., primary keys) of tables at ends of arrows stored as values (i.e., foreign keys) in tables at heads of arrows (e.g., taxonomic information stored in the species table is queried and displayed on factsheets based on species identifier/primary key).

status, information source) and supporting reference documentation (Figure 1). Several accessory tables are also present in the database, primarily used for mapping and data lookup. The full Federal Geographic Data Committee (FGDC) compliant metadata, containing a description of our database schema, can be found at http://1.usa.gov/1uI8cGT.

Occurrence data

Once verified, the reports are entered into the NAS occurrence database, the main component of the NAS database. Each record consists of a species at a place and time, our source for the report, and various other information when available. Each report is georeferenced as precisely as possible using reported geographic coordinates and/or locality description and categorized at several hierarchical levels—nation, state, county, and hydrologic unit code (HUC). As the accuracy of location reporting varies depending on a variety of factors (e.g., age of record, record source), all locations are assigned a qualitative assessment of the accuracy of the georeferenced location: sites are designated as accurate (reasonably close to collection location; e.g., mouth of Smith Creek), approximate (in the general vicinity of the collection; e.g., a pond in Gainesville), or centroid (center of a polygon; e.g., Alachua County or Potomac drainage). Because these organisms are aquatic, they are limited by geological and hydrological boundaries of watersheds. Therefore the NAS Program relates all records to drainages, using the USGS Hydrologic Unit Code (HUC) system (Seaber et al. 1987) and the updated version, the Watershed Boundary Dataset (USDA/NRCS 2011).

QA/QC practices

Prior to entry into the NAS database, records are reviewed to ensure information accuracy and are georeferenced (if no coordinates are included). Our quality assurance process (described in Fuller et al. 2013) varies with the source of the report, the species reported, and the likelihood of occurrence at the reported location. Generally, reports from sources with some degree of training in biology, taxonomy, natural history, or species identification (e.g., scientific publications, museum collection databases, state/federal/tribal natural resource agency personnel) are accepted with minimal review; and more scrutiny is given to news/magazine articles, reports from the general public, and informal web sites (personal sites or discussion forums). Reports from areas
where a nonindigenous species is established and common receive less scrutiny than reports from localities where a species is rare or is novel. Fuller et al. (2013) contains a full description of data assessment practices prior to entry into the NAS database. All supporting documentation for occurrences records (e.g., emails, literature citation, photographs) is stored in electronic formats, and are linked to each record (see ‘Supporting Documentation’ below). Existing records in the NAS database undergo review in several forms to assess accuracy and correct errors. Upon entry into the NAS database, all records are included in a QA/QC list to be checked by a staff member who did not perform initial data entry. Users of the NAS database and website provide near constant review of our publicly-accessible data, and will submit problems or corrections. Species factsheets and distribution maps (including all records for a species) undergo periodic review by regional, topical, or taxonomic experts (e.g., ichthyologists, biologists specializing in Everglades ecology). Much of the data have undergone specific review for other purposes. For example, all fish data were reviewed prior to publication of Fuller et al. (1999), and all data for the southeastern United States were reviewed by the states (South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, and Texas) for an Aquatic Nuisance Species Task Force Gulf and South Atlantic Regional Panel project in 2005.

**Supporting documentation**

All supporting documentation for a record within the NAS database is stored in one of several formats. All reports derived from personal communications or through our online sighting report form include an electronic copy of the communication or form attached directly to the specimen record. Any images supporting a particular specimen record are stored in a separate image database and linked to that record, and are visible on the record’s public view. The image database also contains generic images of each species, not linked to individual specimen records, used on informational factsheets. Copyright clearance is obtained for all photos used on the NAS website. Occurrence reports that are derived from the literature are linked to a reference database. This database currently contains over 15,000 citations: more than 4,600 used as supporting evidence for occurrence records, and the remainder comprising references with unrecorded occurrence data or literature of general interest on invasion biology, ecology, systematics, taxonomy, or species biology. Electronic versions of references are stored, when available, and linked to each reference record. The reference database is open to public query (http://nas.er.usgs.gov/queries/references/default.aspx) by any field in a citation (author, title, journal, etc.), as well as key words. However, documents are not served or distributed in accordance with copyright provisions.

**Information distribution**

Access to public portions of the NAS database is provided online at http://nas.er.usgs.gov. The primary information products of the NAS database are summary factsheets on species’ biology, ecology, and impacts, and records of species occurrences presented both in tabular format and distribution maps. Collection data and distribution maps are updated in real time as new records are entered into the database. Additionally, the NAS web site provides graphs summarizing occurrence data across temporal and geographic ranges, and taxonomic groups.

**Occurrence records**

Each individual specimen record in the NAS database is publicly visible, displaying a variety of information (if available) including: taxonomic identity, date and location of observation, means of introduction and population status in a geographic location, and reference documentation. A map of the occurrence is provided for spatial orientation (Figure 2).

**Alert system**

The NAS alert system was developed in 2004 as a service to distribute information on new introductions to interested users. The system is designed to notify registered users of new occurrences based on their desired categories of interest and are generated at country, state, county, and drainage levels. Timely generation of alerts represents an important component of Early Detection and Rapid Response (EDRR) systems so that monitoring strategies can be prioritized and management actions can be initiated, including the potential of eradication. Fuller et al. (2013) described the system and characterized the alerts and users. An archive of the alerts is located on the website (Figure 3).
Overview and history of the USGS NAS Database

Figure 2. Sample occurrence record from the U.S. Geological Survey’s Nonindigenous Aquatic Species database based on a personal communication.

Figure 3. Alert System archive of the U.S. Geological Survey’s Nonindigenous Aquatic Species database. Reports that are new to the country or to a state, county, or HUC within the past six months to one year are classified as alerts. These are sent to registered users and archived on the Nonindigenous Aquatic Species web site. The archive is searchable based on taxonomic group, species, state, and years.
Fact sheets

Species fact sheets are written summaries of a species' biology, taxonomy, identification, ecology, current known native and introduced ranges, ecological impacts, and management (similar to the species accounts in Fuller et al. 1999). Fact sheets provide photos of nonindigenous species (when available), and a generalized distribution map of both native (for species indigenous to the United States) and/or introduced ranges at the watershed (HUC8) level. Fact sheets exist for the majority of species within the NAS database, and are continually updated and expanded as new information becomes available.

Database searches

The NAS web site provides immediate access to new occurrence records as they are entered through a user-friendly, real-time interface with the NAS database. The NAS website presents users with both map- and text-based search capabilities (Figure 4). Map-based queries allow users to search based on predefined geopolitical or hydrologic boundaries of interest. Simple text-based searches allow for custom queries on scientific and common name; advanced search functionality provides query refinement on establishment status, higher taxonomy, introduction pathway, and year range, or the ability to search for a specific specimen record. Both geographic and textual searches return species lists matching the query parameters, and provides links to factsheets, collection information, and distribution maps for each individual species. Most users obtain the information they need from our on-line tools; however, we provide custom queries and datasets for people with more specific needs.

Distribution maps

One of the more visible portions of the NAS database and website is the species distribution maps. In particular, NAS maps of zebra and quagga mussel [Dreissena polymorpha (Pallas, 1771) and D. rostriformis bugensis Andrusov, 1897], Asian carps [i.e., silver carp (Hypophthalmichthys molitrix (Valenciennes in Cuvier and Valenciennes, 1844)), Bighead Carp (H. nobilis (Richardson, 1845)), Black Carp (Mylopharyngodon piceus (Richardson, 1846))], Lionfish (Pterois miles and P. volitans); and Asian Tiger Shrimp (Penaeus monodon) have appeared in many publications, invasive species presentations, and news outlets. Prior to 2008, the program mapped all distributions at the watershed scale (both HUC6 and HUC8). Many records, particularly those from earlier publications, lack precisely documented geographic coordinates for the location of introduction. Beginning in 2008, the NAS program began an initiative to geo-reference all records in the database based on site descriptions and through the use of georeferencing tools and databases including GEOLocate (Rios and Bart 2010) and the USGS Geographic Names Information System (U.S. Board on Geographic Names 2014). All locations in the NAS database receive a qualitative assessment of the spatial accuracy of the assigned geographic coordinates (described above).
Georeferencing of all historic and future data allowed for the generation of point distribution maps, which were launched in November 2009. Distribution maps are available in two (user-selectable) formats: cluster maps and status maps (Figure 5). Cluster maps (the default) groups adjacent records into larger clusters, with marker size and color indicating the total number of records contained within a cluster. Status maps present all records simultaneously, with current population status indicated by marker color and spatial accuracy by marker shape. Both map types link each point back to full specimen records and display native ranges for North American taxa, derived from a variety of sources including primary scientific literature, state natural history books and databases, NatureServe (NatureServe 2014), and museum specimen information. Maps are customizable by allowing the user to choose various backgrounds, as well as political and hydrological outlines. In 2013 animated maps were made available to allow the users to visualize spread of species.

Database overview

The dataset used in this characterization contains 268,771 specimen occurrence records from 1,051 unique species. Although the NAS database contains specimen records dating back to 1800, the majority of records are documentation of introductions within the last ~40 years (Figure 6). Increased numbers of records dated within the last ~40 years are at least partially a result of an increased awareness of impacts associated with nonindigenous species and the growth of research in invasion biology, increasing rates of communication of nonindigenous species occurrences, or other factors. Similarly, a low number of specimen records dating prior to ~1900 could be due difficulty in documenting historical records of species introductions and observations of nonindigenous species (e.g., no available documentation or problems accessing literature sources; missing or ambiguous species or locality descriptions). However, the increased rate of species introductions is corroborated by other sources (Courtenay et al. 1991), and is consistent with trends in freshwater and marine environments (Mills et al. 1993; Mills et al. 1996; Cohen and Carlton 1998).

Temporal accuracy of NAS specimen records

Figure 7 shows the temporal accuracy assigned to records within the NAS database. A majority of the specimen records within the database include the actual year of observation/introduction (i.e., when a species was collected or reported), including most records for observations within the last 50 years. Publication of state fish books (e.g., Rohde et al. 2009) and fish stocking information released by state agencies result in spikes of temporal accuracy equal to “Publication Year”.

Geographic accuracy of NAS specimen records

Most records within the NAS database are assigned to a specific geographic location (designated as accurate), due to either direct reporting of geographic coordinates or a precise description of the collection location. Some records are more vague, as described above, and are classified as approximate or centroid for location accuracy. More than 70% of the locations are specific enough to be classified as accurate (Figure 8).
Figure 5. Examples of the two types of maps served by U.S. Geological Survey’s Nonindigenous Aquatic Species database. The cluster map (Largemouth Bass, *Micropterus salmoides*) which indicated the number of records for a given area; and the population status map (Bighead Carp, *Hypophthalmichthys nobilis*) which displays individuals records and indicates both population status and geographic accuracy of reports.

References

The NAS reference database comprises all literature sources (including books, news stories, articles from scientific journals, agency reports, etc.) that are used as documentation for specimen records, as well as references containing supplementary information used to create species-specific fact sheets. The NAS reference database includes 14,942 references published between 1800–2013, of which 4,668 are used for documentation of specimen occurrence records. A large increase in the number of references after ~1990 (Figure 9) is due to the increasing importance and use of internet-associated sources (e.g., news websites, online state agency stocking reports, museum collection databases) of species occurrence data, and the increase of research efforts in invasion biology.
Overview and history of the USGS NAS Database

Figure 6. Total record count per decade of unique species (dark grey circles) and individual specimen records (light grey squares) present in the U.S. Geological Survey’s Nonindigenous Aquatic Species database with an introduction year between 1800–2013.

Figure 7. Temporal accuracy of specimen records present in the U.S. Geological Survey’s Nonindigenous Aquatic Species database with an introduction year between 1800–2013. Counts are total over each decade. ‘Actual’ indicates specific year of observation was given for a record; ‘Publication Year’ indicates that no year of observation was given and publication year of record documentation was substituted; ‘Estimated’ indicates that the observation date is unknown but an approximate range was present in the record documentation.

Figure 8. Spatial accuracy of specimen records present in the U.S. Geological Survey’s Nonindigenous Aquatic Species database with an introduction year between 1800–2013. ‘Accurate’ indicates a unique site description or geographic coordinates contained within occurrence documentation; ‘Approximate’ indicates a potentially ambiguous site description or a general location given in occurrence documentation; ‘Centroid’ indicates coordinates are the centerpoint of a specific geographic polygon (i.e., administrative or watershed boundary) calculated by GIS; ‘Other’ indicates specimen records that do not fall into any other category (primarily historic or extremely ambiguous records).

Figure 9. Number of references used as occurrence record documentation within the U.S. Geological Survey’s Nonindigenous Aquatic Species database.
Uses and products

Occurrence data stored in the NAS database has been used to examine taxonomic, geographic, or temporal influences on nonindigenous species’ distributions, dispersal pathways and vectors, introduction sources, or establishment success. Data from the NAS database has been used in: conducting risk analyses (Herborg et al. 2007; Jenkins et al. 2007; Whittier et al. 2008; Zajicek et al. 2008), preparing field guides for early detection (Schofield et al. 2005; Schofield et al. 2009), predictive modeling (Drake and Lodge 2006; Mercado-Silva et al. 2006; Bossenbroek et al. 2007; Chen et al. 2007; DeVaney et al. 2009), state aquatic nuisance species management plans (California Department of Fish and Game 2007; Idaho Invasive Species Council Technical Committee 2007; South Carolina Aquatic Invasive Species Task Force and South Carolina Department of Natural Resources 2007), species-specific management plans (Western Regional Panel 2009), regional management (Rodgers et al. 2010), Congressional testimony (Thayer 2010), national assessments (Rahel 2000; Heinz Center 2002, 2008; Stohlgren et al. 2006), to document species invasion (Schofield 2009; Fuller et al. 2014), and national policy making (silver carp listing in the Lacey Act) (U.S. Fish and Wildlife Service 2007). The NAS database is referenced in the National Invasive Species Management Plans (National Invasive Species Council 2001, 2008) and the Aquatic Nuisance Species Task Force Strategic Plan (Aquatic Nuisance Species Task Force 2007) as a system to be supported and used as a national repository of data. A larger list of works that have used or cited the NAS database and website is presented in Appendix 1. Several large scale analyses of the NAS database have also been published. Fuller et al. (1999) summarized 20 years of data contained in the NAS database on introduced fishes in the United States, and its species accounts were the genesis for the species factsheets present on the NAS website. Fuller et al. (2013) analyzed various components of the NAS alert system (e.g., subscriber demographics; taxonomic, spatial, and temporal trends in alert composition). Stohlgren et al. (2006), compared spatial richness patterns in various introduced species nationwide. Guo and Olden (2014) analyzed spatial scaling of nonindigenous fish richness across the U.S.

Partnerships

The NAS program works closely with other state and federal natural resource agencies and develops special tools with partnerships, such as integrated reporting and filtered website views. Working with the Smithsonian Environmental Research Center (SERC) Marine Invasions Lab, the NAS program has helped to create an integrated data system called NISbase (Stevens et al. 2003) that provides a unified search interface to both the NAS database and the SERC database on nonindigenous estuarine and marine species. Through other partnerships, NAS offers customized, regional views of the database that allow managers to focus on a specific area of interest, including the NOAA GLANSIS website (GLANSIS 2014) and a regional view of the Columbia River Basin provided for the 100th Meridian Initiative-Columbia River Basin Team (CRBAIS 2011). The NAS database is also a member of the Global Invasive Species Information Network (GISIN 2014), the Global Biodiversity Information Facility (GBIF 2014), Biological Information Serving Our Nation (BISON 2014), and Ocean Biogeographic Information System (OBIS 2014) and provides data to these larger networks.

Additionally, the NAS program has ongoing data sharing agreements with other nonindigenous species database programs (EDDMapS 2014; iMapInvasives 2014).

Acknowledgements

The NAS program and staff thank state and federal aquatic nuisance species coordinators, as well as many other biologists and the general public for reporting these occurrences to us. We also thank researchers for donating data sets and our past and present funders including the U.S. Geological Survey, U.S. Fish and Wildlife Service, U.S. Environmental Protection Agency, and National Oceanic and Atmospheric Administration. Use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References


(Accessed 11 December 2014)


Overview and history of the USGS NAS Database


Rohde FC, Arndt RG, Foltz JW, Quattro JM (2009) Freshwater fishes of South Carolina. University of South Carolina Press, 430 pp


