

---

# Authoritative Lists of Non-Native Species for AK, HI, L48

*A Data Management Plan created using DMPTool*

Creators: Annie Simpson, Elizabeth Sellers, meyer@usgs.gov

Affiliation: U.S. Geological Survey (USGS)

Template: U.S. Geological Survey (USGS)

ORCID iD: <http://orcid.org/0000-0001-8338-5134>

## Project abstract:

Invasive species are a subset of non-native (or alien) species, and knowing what species are non-native to a region is a first step to managing invasive species. People have been compiling non-native and invasive species lists ever since these species started causing harm, yet national non-native species lists are neither universal, nor common. Non-native species lists serve diverse purposes: watch lists for preventing invasions, inventory and monitoring lists for research and modeling, regulatory lists for species control, and nonregulatory lists for raising awareness. This diversity of purpose and the lists' variation in geographic scope make compiling comprehensive lists of established (or naturalized) species for large regions difficult. However, listing what species are non-native in an area helps measure Essential Biodiversity Variables for invasive species monitoring and mount an effective response to established non-native species. In total, 1,166 authoritative sources were reviewed to compile the first comprehensive non-native species list for three large regions of the United States: Alaska, Hawaii, and the conterminous United States (lower 48 States). As of July 2018, the list contains 11,344 unique names: 598 taxa for Alaska, 5,848 taxa for Hawaii, and 6,675 taxa for the conterminous United States. The list is available to the public from ScienceBase (<https://doi.org/10.5066/P9E5K160>), and the intent, though not a guarantee, is to update the list as non-native species become established in, or are eliminated from, the United States. The list has been used to annotate non-native species occurrence records in the U.S. Geological Survey all-species mapping application, Biodiversity Information Serving Our Nation (BISON, <https://bison.usgs.gov>). A USGS Open-File Report discusses creation and analysis of the non-native list (<https://doi.org/10.3133/ofr20181156>).

Last modified: 10-31-2018

## Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

# Authoritative Lists of Non-Native Species for AK, HI, L48

---

## Product Information

First comprehensive list of non-native species established in three major regions of the United States of America

A compilation of authoritative assertions of non-native (alien) status for species established within one of three geographic areas: Alaska, Hawaii, and the conterminous United States herein referred to as the lower 48 States and abbreviated as L48. The data product consists of three species sublists and associated information (compiled into a master list), plus a bibliography and digital copies, where possible, of the authorities asserting the species' nonnative status in these locations. Also, to assist with ongoing updates to the non-native species list, there is an ongoing and parallel informal compilation of watchlists of species likely to occur in the three regions in the future.

Annie Simpson, biologist and information scientist  
U.S. Geological Survey, Mailstop 302  
12201 Sunrise Valley Drive  
Reston VA 20192  
asimpson@usgs.gov  
+1-703-648-4281

### Co-authors:

Annie Simpson, biologist and information scientist  
U.S. Geological Survey, Mailstop 302  
12201 Sunrise Valley Drive  
Reston VA 20192  
asimpson@usgs.gov  
+1-703-648-4281

Meghan C. Eyler (nee Mannas), environmental information specialist  
U.S. Geological Survey  
11510 American Holly Drive  
Laurel, Maryland 20708 USA  
meyler@usgs.gov  
+1-301-497-5882

## Version 7, 2018-10-31

### Acquire Stage

Data acquisition determined by product owner and team; data acquisition is facilitated by small amounts of project funding to hire data collectors; data acquisition is also opportunistic and resource-dependent.

The data providers for the product are USGS; Other DOI agencies; Other Federal agencies outside DOI; State/local agencies; NGOs; Academic/research institutions; and Personal/individual contributors. Note that data providers are authorities making an assertion as to the non-native status of one or more species in one or more regions. Providers are deemed reliable and acceptable if they come from one of the above types of sources, include contact information, and are generally considered to have expertise in the area of non-native species science.

Data are species lists that are publicly available and all authorities are referenced/cited.

Data types include: Excel spreadsheets; Google spreadsheets; Databases of different formats that are converted to spreadsheets or simple text; MS Word files; PDF; Other (html).

Data are accepted by email attachment, FTP site, and found through online and library search.

Local network drive, DOI Google Drive.

A preversion of the list is deposited with the Knowledge Network for Biocomplexity, [doi:10.5063/F1RF5S79](https://doi.org/10.5063/F1RF5S79).

The definitive version of the list, an FGDC metadata record, other supporting documents, and all updated versions are deposited with USGS' ScienceBase:

Simpson, Annie, Meghan C. Eyler, Matthew Cannister, Reko Libby, Natasha Kozlowski, Elizabeth Sellers, and Gerald F. Guala, 2018, Dataset of the first comprehensive list of non-native species established in three major regions of the United States: U.S. Geological Survey data release, <https://doi.org/10.5066/P9E5K160>.

A USGS Open-File Report discussing the list is available here:

Simpson, Annie, and Eyler, Meghan C., 2018, First comprehensive list of non-native species established in the United States, segregated by major regions: U.S. Geological Survey Open-File Report 2018–1156, 15 p., <https://doi.org/10.3133/ofr20181156>.

Data and citation standards follow the Integrated Taxonomic Information System (ITIS, <https://www.itis.gov/standard.html>). Where available, authoritative source DOIs are included.

Metadata from sources is a citation that follows the ITIS citation format. Minimal information required: author, year, title, and link. If source is not stable, digital copy of the authoritative source is saved to disk.

N/A

Completeness of data acquisition is determined by data team. First draft has been completed. Final version anticipated by the end of the 2018 calendar year, to be updated in an ongoing manner as resources allow.

## Process Stage

- Yes, we do perform data processing that cover the Process Stage

Data received is analyzed for suitability and processed to fit the format of the list, which is made available in .csv and .xlsx formats.

More than a thousand separate authorities of many disparate types were consulted to create the non-native species lists for the US over a six year period. Within the protocol to update the lists in a reliable ongoing manner, updating each type of authority will require a somewhat different approach.

-----

## DESCRIPTION OF PROCESSING WORKFLOW

### Resource Type | Publication Regularity | Publication Frequency | Planned update methodology

**Article** | Regular | Weekly, Monthly, Quarterly, Annually | Review journal as published; monthly search of Google Scholar using key terms

**Book** | Irregular | Once | Monthly Web & library search for new sources

**Brochure** | Irregular | Once | Quarterly Web & library search for new sources

**Circular** | Irregular | Ongoing | Quarterly web and library search for new sources

**Database** | Irregular | Ongoing | Biennial review for updates

**Environmental Assessment** | Irregular | Once | Biennial review of organization's publications

**Technical Report** | Irregular | Once | Biennial review of organization's publications

**Thesis** | Irregular | Once | Annual review of Google Books, PQDT Open, OATD.org, & NDLTD.

**Watch list** | Irregular | Ongoing | Add names to the watch list(s) as they are opportunistically discovered. Perform a web search with names from the watch lists biennially to determine if names should be moved to non-native list(s).

**Website** | Irregular | Ongoing | Biennial review of existing authoritative Websites & Web search for additional authoritative Websites

-----

In addition to the ongoing search regimen of reviewing the existing authoritative sources and email alerts from relevant journals, general web searches for the watch list species names will be conducted to keep the non-native species list current.

Until there is a web service allowing automatic uploads of invasive species lists by stakeholders, direct suggestions for additions to any of the non-native species lists will gratefully be received by the authors (and should include an authoritative source reference).

It is planned that the non-native species list, as it is updated with new information (including corrections), will be labeled with a version number and emission date.

Source information is generally available in .txt, .xlsx, .doc, .html, or .pdf. It is converted to .xlsx by copy/paste or occasionally (especially if .pdf) by retyping.

The data fields, which are still somewhat in flux, are as follows:

Region	scientificName	Kingdom	Phylum/Division	Class	Order	Family	providedSciName	Authority	TSN	Common Name
websiteURL	introduced or invasive	Source Type	Source	Version	Acquisition Date	Source Comment	Original description indicator	Reference	Author	Title
Publication Name	Listed	Publication Date	Actual	Publication Date	Publisher	Publication Place	ISBN	ISSN	Pages	Publication Comment

The required fields vary somewhat depending on the source, but at a minimum include:

Region scientificName Authority Source Type Source Acquisition Date.

The data team tasks related to incoming data are variable, and depend on the format and size of the new data.

Incoming data is error checked. Scientific names are validated in the Integrated Taxonomic Information System (ITIS). Other data fields are assumed correct. The validity of the assertion of non-native status itself may subsequently be questioned and the entry deleted or moved to a watch list if: the scientific name is not in ITIS and not reaffirmed by other authoritative taxonomic sources; other authoritative sources state the species is not present; typographical errors in scientific names are found; entries are found to be duplicate or synonymous taxonomic names.

Microsoft Word to assist in formatting data.

Microsoft Excel to hold the data and sort and display it.

ITIS to validate taxonomic names.

Google Sheets to store the data.

Adobe Pro to extract text from image-based pdfs.

Various browsers (MS Explorer, Google Chrome, Mozilla Firefox), species databases, and search engines to discover authoritative nonnative species lists for AK, HI, and L48.

Our team tracks the provenance of the dataset as any modifications, transformations, edits, or decisions to accept or reject data points are made. Provenance is essential to the dataset, which consists of assertions of the presence and non-native status of a species in an area.

This tracking involves the inclusion of an authority or data source with each assertion that a species is non-native in a given area. The actual authorities' lists and manuscripts themselves are also saved digitally as written proof of the individual assertions and to justify, as necessary, each species' inclusion in the list.

The dataset is maintained in Google Sheets and in MS Excel and the 'save as' function is used at regular intervals to maintain a version of the dataset that can be 'rolled back' to previous versions, if necessary. Past versions of the dataset are maintained but not publicly shared.

There are no access constraints to the dataset.

## Analyze Stage

- Yes, we perform analyses which will cover the Analyze Stage

A publication is in press which analyzes the data:

Simpson, Annie and Eyler, Meghan C., 2018, First comprehensive list of non-native species established in the United States, segregated by major regions: U.S. Geological Survey Open-File Report 2018–1156, 16 p., <https://doi.org/10.3133/ofr20181156>.

Expertise includes: Completion of at least a Bachelor's degree (graduate degree preferred) in appropriately related field in life sciences and/or library science; Familiarity with scientific taxonomic and non-native species concepts and fields of research; High proficiency and experience in Web- and library-based searches for authoritative biological species lists; Basic biological and data management knowledge; Attention to detail; Experience with Microsoft Excel and the ability to perform lookups and cull duplicate names from the list; Knowledge and use of metadata; Understanding of the difference between primary and secondary references; Ability to read and critically assess scientific and technical literature; Highly organized and well-versed in effective file/records and information management.

Training includes: EndNote; Detailed find/replace in Word; Advanced filtering techniques in Excel; Use of Integrated Taxonomic Information System and associated tools for scientific name assessment and validation.

Methods used to create the integrated dataset have varied somewhat over time. Additional fields have been added from those that were used at the start of the project. Initially, each region (AK, HI, L48) had a lead data searcher and the lists were joined afterwards within one Excel spreadsheet. The interpretation of the content of the citation fields follow the rules established by ITIS and described here: [http://www.itis.gov/submit\\_guidelines.html](http://www.itis.gov/submit_guidelines.html)

Each lead data searcher was trained in the methods to discern an authoritative source; how to extract scientific names from the source; how to generate from the source and other reference databases such as ITIS all of the information in the spreadsheet columns, specifically:

**Region** [Required; Alaska (AK), Hawaii (HI), and/or contiguous United States (L48)]

**scientificName** [Required; an accepted genus, species, subspecies, variety, or hybrid name, preferably validated in ITIS. Genus names only allowed if all members of the genus are non-native to the area in question.]

**Kingdom** [Required; one of seven names as specified in ITIS or another naming authority (Bacteria, Archaea, Protozoa, Chromista, Plantae, Fungi, Animalia)]

**Phylum/Division** [Optional; as provided by the taxonomic authority]

**Class** [Optional; as provided by the taxonomic authority]

**Order** [Optional; as provided by the taxonomic authority]

**Family** [Optional; as provided by the taxonomic authority]

**providedSciName** [Optional; only included if the scientific name provided by the authority is an unaccepted synonym of a valid/accepted name in a taxonomic authority.]

**Authority** [Required; the abbreviated name for the resource asserting the taxa's non-native species status. Generally, but not always, followed by (year); occasionally also includes a letter to make the reference unique.]

**TSN** [Optional; unique Taxonomic Serial Number for the taxa, as provided by the Integrated Taxonomic Information System, <https://www.itis.gov>]

**Common Name** [Optional; provided by ITIS, the authority, or another reliable source; may be a very generic term.]

**websiteURL** [Optional; provides a link back to an online authority; occasionally may not resolve to the resource.]

**introduced or invasive** [Required; an assertion provided by or derived from the authority describing the non-native or invasive status of the taxa.]

**Source Type** [Required; loosely controlled vocabulary consisting of the following: Article; Book; Brochure; Chapter; Circular; Database; Ecological Risk Screening; Environmental Assessment; Guidebook; Handbook; Manual; Pest Advisory; Restoration Plan; Technical Report; Thesis; Website]

Assertions of non-native status for a species are accepted if the authority is a trusted source, such as a governmental or governmental-affiliated, non-governmental organization, international biodiversity organization, academic institution, or a taxonomic expert. If a list is found on the Web and judged as authoritative, it is downloaded for future reference. Authorities, in addition to being cited within each record, are accumulated in a citation spreadsheet. These criteria are part of our documentation.

Our workflow is described in a USGS Open File Report that accompanies the non-native species list. The assumptions made were that sources are reliable and scientific names could be validated. There were no estimates of uncertainty, tools algorithms, models, or scripts used within our workflow.

There is an FGDC metadata record for the dataset available through ScienceBase at <https://doi.org/10.5066/P9E5K160>.

## Backup and Secure

Acquire, Process, Analyze, Preserve: Matthew Cannister, Meghan Eyler, Gerald Guala, Natasha Kozlowski, Reko Libby, Elizabeth Sellers, Annie Simpson.

Publish/Share: Meghan Eyler and Annie Simpson.

Originally, each regional dataset was generated and maintained by regional lead data searcher of the data. These were merged into a master list in 2016. One shared Google Drive master list copy of the data is currently worked on by one person at a time (with a document title name change notice of being checked out). Pieces of the dataset may also be extracted and worked on individually, then merged back to the master list. There is no simultaneous work performed on the master list.

Operational copies are stored on individual computers or worked on in Google Drive. When working subsets of the master list are extracted for work on an individual's computer, work on the master list is 'frozen' for that part of the data until it is again uploaded and added to the master list.

The operational copy is backed up on an external drive several times a day by an automated process, and also emailed between team members as an attachment. Annie Simpson bears primary responsibility for backups.

Risks are low for permanent data loss, and virtually impossible for extensive loss because of the large number of historic copies stored by the team in many locations. The dataset is not overly large, making this possible.

## Preserve Stage

The dataset is offered within BISON's Solr web service. USGS has published a peer-reviewed USGS Open File Report. The dataset and its metadata is stored in ScienceBase, and the list has been integrated into the Biodiversity Information Serving Our Nation (BISON) species mapping application on the Web, <https://bison.usgs.gov>.

The dataset is nonproprietary and includes freely available information whose authors are cited.

Our data package does not include sensitive or restricted data and will be openly accessible.

Raw data, in the form of lists/assertions by authorities of a species' nonnative status, are preserved on individual computers and cited using the ITIS citation format.

All documentation (standards, methods, processes, scripts, code, models, graphs) related to processing and analysis steps have been preserved so that an external user can trace workflows to understand conclusions and interpretations reached. Now completed and peer reviewed, the dataset has been made available in tab delimited format and assigned a DOI on ScienceBase. An FGDC metadata

record, supporting documents, and a readme file are included. See: <https://doi.org/10.5066/P9E5K160>.

Analysis of final data package is preserved within a peer reviewed USGS Open File Report, on the USGS pubs warehouse, and within ScienceBase. The dataset is posted on ScienceBase and the Knowledge Network for Biocomplexity and periodically updated on ScienceBase.

ScienceBase: <https://doi.org/10.5066/P9E5K160>

Knowledge Network for Biocomplexity: <https://doi:10.5063/F1RF5S79>

USGS pubs warehouse: <https://doi.org/10.3133/ofr20181156>

## **Publish/Share Stage**

The metadata will be included in the data packages, as well as described in a USGS Open File Report.

There will be no restrictions for end-user access.

ScienceBase has assigned this DOI to the dataset:

<https://doi.org/10.5066/P9E5K160>

and has this landing page:

<https://www.sciencebase.gov/catalog/item/5b911a5ce4b0702d0e808588>

The Knowledge Network for Biocomplexity also has assigned this DOI to the dataset:

[doi:10.5063/F1RF5S79](https://doi:10.5063/F1RF5S79).

The USGS pubs warehouse has assigned this DOI to the Open File Report <https://doi.org/10.3133/ofr20181156>.

The data will be distributed by Web site, offered through web services, included in a peer-reviewed USGS Open File Report, and has been incorporated into the BISON application.

The data is searchable in Solr through the BISON web service. Additional Web services to provide this information through the BISON project's online platform are also planned in order to share the non-native species list widely.

Discoverability will be improved by optimizing discoverability in Google; optimizing discoverability in USGS Science Data Catalog; and including it in a published peer-reviewed USGS Open File Report.