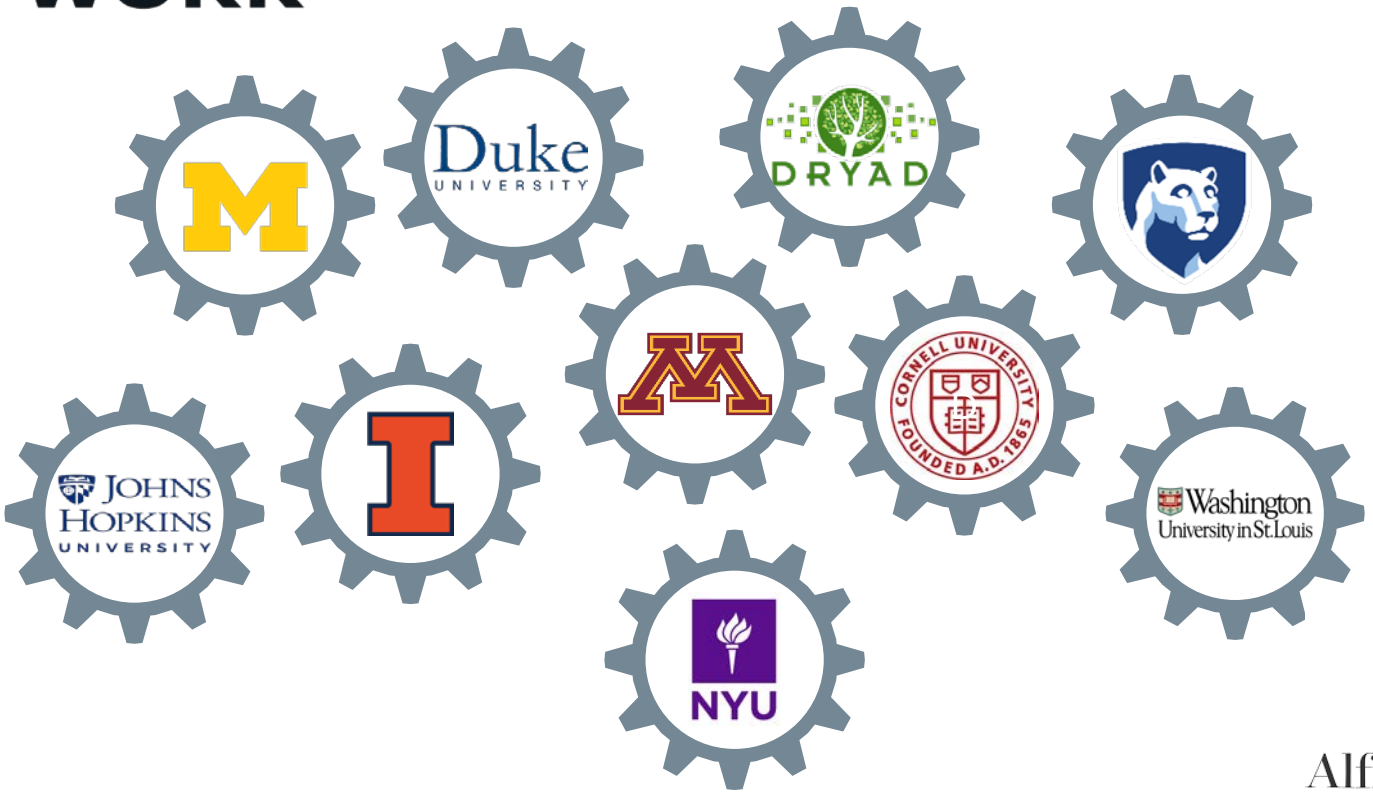
The Data Curation Network

Cynthia Hudson Vitale
Pennsylvania State University

Data Curation Network All Hands Meeting 2019

DATA CURATION NETWORK



Alfred P. Sloan
FOUNDATION

DATA
CURATION
NETWORK

Mission

“The Data Curation Network will enable researchers that are faced with a growing number of requirements to ethically share their research data in ways that make it findable, accessible, interoperable and reusable (FAIR).”

What is data curation?

Data curators enrich research data publications and ensure the data are FAIR, by:

- Finding and adding missing files and documentation
- Screening for privacy disclosure risk
- Detecting and fixing code and other quality assurance issues
- Transforming file formats for long term access
- Arranging and describing files
- Reviewing and augmenting metadata

LIBRARIES
digital conservancy

Search Browse Collections Help Sign in

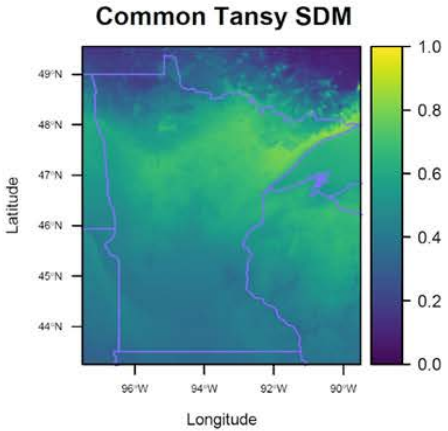
Data Repository for U of M

CORE TRUST SEAL

University Digital Conservancy Home / University of Minnesota / Data Repository for U of M (DRUM) / View Item

Species Distribution Models and Joint Species Distribution Models of Nine Invasive Species in North America

Lake, Thomas, A; Briscoe Runquist, Ryan, D; Moeller, David, A (2019)



Persistent link to this item
https://doi.org/10.13020/z71w-jx69
http://hdl.handle.net/11299/206482

Services
Full Metadata (xml)
View Usage Statistics

Title
Species Distribution Models and Joint Species Distribution Models of Nine Invasive Species in North America

Published Date
2019-08-27

Authors
Lake, Thomas, A
Briscoe Runquist, Ryan, D
Moeller, David, A

Group
Moeller Lab

Author Contact
Briscoe Runquist, Ryan, D (rbriscoe@umn.edu)

Type
Dataset

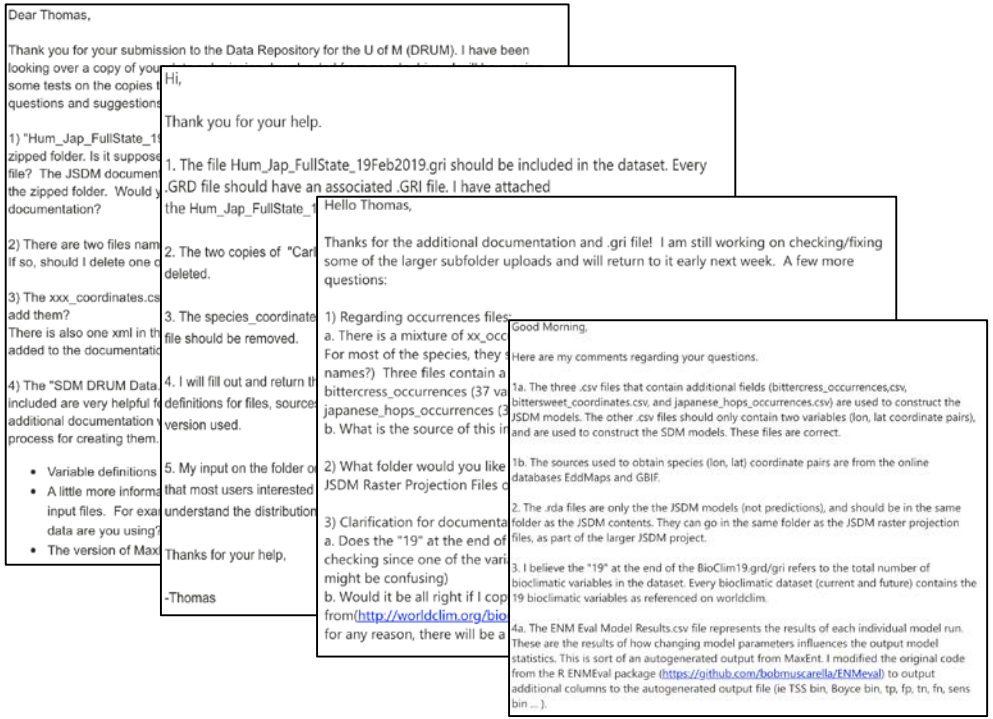
Abstract
Species Distribution Models (SDM) and Joint Species Distribution Models (JSDM) for nine invasive species in North America. Species SDMs include Brown Knapweed (Centaurea jacea), Black Swallowwort (Cynanchum louiseae), Common Tansy (Tanacetum vulgare), Common Teasel (Dipsacus fullonum), Wild Parsnip (Pastinaca sativa), and Dalmatian Toadflax (Linaria dalmatica). Species JSDMs include Japanese Hops (Humulus japonicus), Oriental Bittersweet (Celastrus orbiculatus), and Narrowleaf Bittercress (Cardamine impatiens). These models aim to predict the current and future habitat suitability of nine invasive species in North America. Models were constructed with the

Species Distribution Models and Joint Species Distribution Models of Nine Invasive Species in North America

Lake, Thomas, A; Briscoe Runquist, Ryan, D; Moeller, David, A (2019)

Data Curator: Melinda Kernik
Spatial Data Analyst and Curator, University of Minnesota Libraries and U-spatial

- Identified 4 missing files and 2 unnecessary files
- Created folder organization and hierarchy
- Clarified inconsistent file naming and variable naming conventions
- Added citations for data reuse from EddMaps and GBIF
- Developed a data dictionary for variable names



Dear Thomas,

Thank you for your submission to the Data Repository for the U of M (DRUM). I have been looking over a copy of you some tests on the copies t questions and suggestions

1) "Hum_Jap_FullState_1 zipped folder. Is it suppose file? The JSDM document the zipped folder. Would y documentation?

2) There are two files nam If so, should I delete one c

3) The xxx_coordinates.cs add them?
There is also one xml in th added to the documentatic

4) The "SDM DRUM Data included are very helpful f additional documentation v process for creating them.

- Variable definitions
- A little more informa input files. For exa data are you using?
- The version of Max

Hi,

Thank you for your help.

1. The file Hum_Jap_FullState_19Feb2019.gri should be included in the dataset. Every .GRD file should have an associated .GRI file. I have attached the Hum_Jap_FullState_1

2. The two copies of "Carl deleted.

3. The species_coordinate file should be removed.

4. I will fill out and return th definitions for files, source version used.

5. My input on the folder o that most users interested understand the distribution

Thanks for your help,

-Thomas

Hello Thomas,

Thanks for the additional documentation and .gri file! I am still working on checking/fixing some of the larger subfolder uploads and will return to it early next week. A few more questions:

1) Regarding occurrences files a. There is a mixture of xx_occ For most of the species, they names?) Three files contain a bittercress_occurrences (37 va japanese_hops_occurrences (3 b. What is the source of this in

2) What folder would you like JSDM Raster Projection Files c

3) Clarification for documenta a. Does the "19" at the end of checking since one of the vari might be confusing)
b. Would it be all right if I cop from(http://worldclim.org/bio for any reason, there will be a

Good Morning,

Here are my comments regarding your questions.

1a. The three .csv files that contain additional fields (bittercress_occurrences.csv, bittersweet_coordinates.csv, and japanese_hops_occurrences.csv) are used to construct the JSDM models. The other .csv files should only contain two variables (lon, lat coordinate pairs), and are used to construct the SDM models. These files are correct.

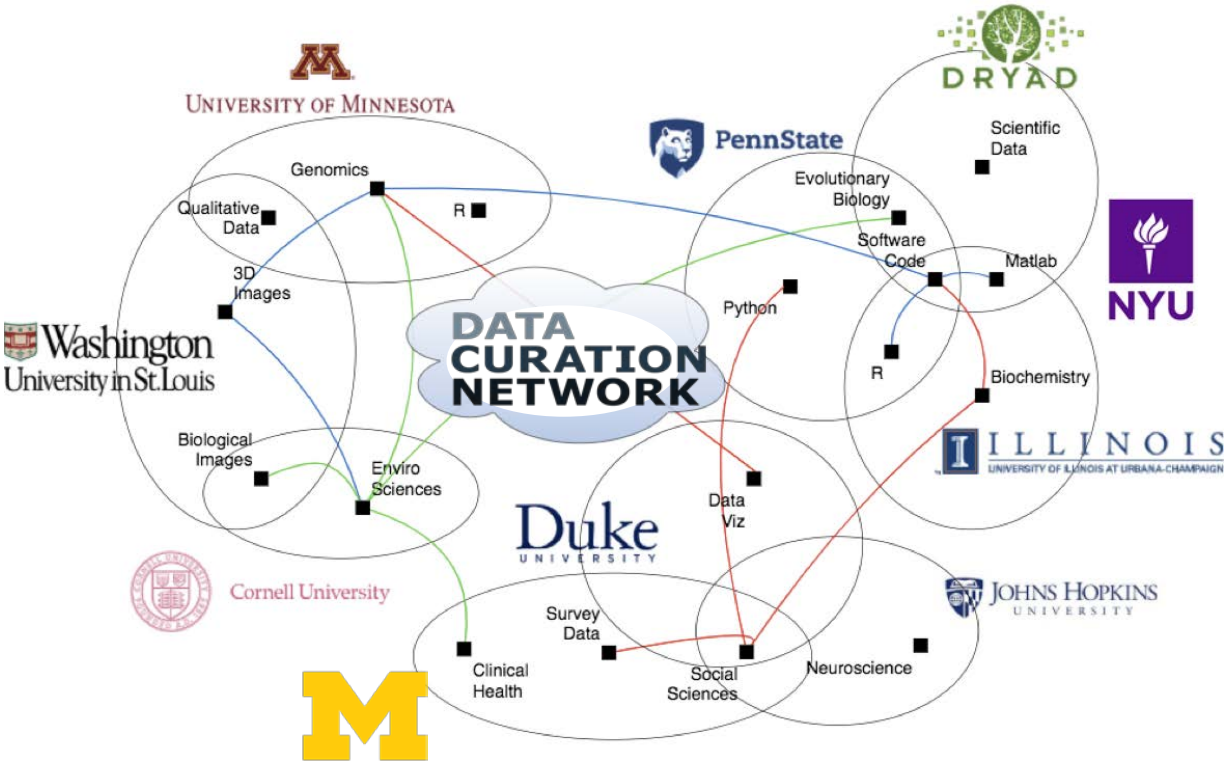
1b. The sources used to obtain species (lon, lat) coordinate pairs are from the online databases EddMaps and GBIF.

2. The .rda files are only the the JSDM models (not predictions), and should be in the same folder as the JSDM contents. They can go in the same folder as the JSDM raster projection files, as part of the larger JSDM project.

3. I believe the "19" at the end of the BioClim19.grd/gri refers to the total number of bioclimatic variables in the dataset. Every bioclimatic dataset (current and future) contains the 19 bioclimatic variables as referenced on worldclim.

4a. The ENM Eval Model Results.csv file represents the results of each individual model run. These are the results of how changing model parameters influences the output model statistics. This is sort of an autogenerated output from MaxEnt. I modified the original code from the R ENMEval package (https://github.com/bobmuscarella/ENMeval) to output additional columns to the autogenerated output file (ie TSS bin, Boyce bin, tp, fp, tn, fn, sens bin ...).

DCN Expert Network



DCN CURATE Steps

DCN Curators will take **CURATE** steps for each data set, that ir

- **C** **Check** data files and read documentation
- **U** **Understand** the data (try to), if not...
- **R** **Request** missing information or changes
- **A** **Augment** the submission with metadata for findability
- **T** **Transform** file formats for reuse and long-term preservation
- **E** **Evaluate** and rate the overall submission for FAIRness.

Table A1. Draft checklist of DCN CURATE steps and FAIRness scorecard

CURATE Actions	Curation Checklist
Check data files and read documentation ▪ Review the content of the data files (e.g., open and run the files or code). ▪ Verify all metadata provided by the author and review the available documentation.	☐ Files open as expected ☐ Issues _____ ☐ Code runs as expected ☐ Produces minor errors ☐ Does not run and/or produces many errors ☐ Metadata quality is rich, accurate, and complete ☐ Metadata has issues _____ ☐ Documentation Type (*circle*) Readme / Codebook / Data Dictionary / Other: _____ ☐ Missing/None ☐ Needs work
Understand the data (or try to) ▪ Check for quality assurance and usability issues such as missing	*Varies based on file formats and subject domain. For example….*

DCN Growth and Sustainability

- Curated 50 data sets since Jan 1, 2019!
- 2 new members in Year 2
- Aim to add two more in Year 3
- Canada and Dutch groups planning stages to launch their own network
- Exploring fiscal and administrative models to support beyond grant