

You Reap What You Sow: Understanding Data Quality in the NPDN Repository

Matthew A. Bertone¹, Clarissa J. Balbalian², Karen K. Rane³, Erin C. Hill⁴, Patrick J. Shiel⁵, Eleanor Lopez⁶, Mike Hill⁷, Tim Deppen⁸, Tania Brenes-Arguedas⁹, and Tom Creswell¹⁰

1. Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA matt_bertone@ncsu.edu; 2. Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University, Mississippi State, MS, USA c.balbalian@msstate.edu; 3. Department of Entomology and Department of Plant Sciences and Landscape Architecture, University of Maryland, College Park, MD, USA kraner@umd.edu; 4. Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, WUSA erinhill12@msu.edu; 5. Domestic and Emergency Scientific Support, USDA-APHIS-PPQ Science and Technology, Raleigh, NC, USA Patrick.J.Shiel@usda.gov; 6. Soil, Plant and Pest Center, Nashville, TN, USA, University of Tennessee, elopez@utk.edu; 7. Center for Environmental and Regulatory Information Systems, Purdue University, West Lafayette, IN, USA, mikehill@purdue.edu; 8. Center for Environmental and Regulatory Information Systems, Purdue University, West Lafayette, IN, USA, tdeppen@purdue.edu; 9. University of California, Davis, CA, USA, tbrenesa@ucdavis.edu; 10. Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA, tcreswell@purdue.edu

ABSTRACT

The quality and consistency of the data in the NPDN National Data Repository (NDR) relies on clear entry guidelines and terminology. The NPDN National Data Committee provides these guidelines and standardizes terms used in the NDR. The committee also reviews nomenclature for taxonomic accuracy, and evaluates new dictionary entry and data requests. In recent years the committee has identified and begun to address several issues regarding data in the NDR and how users can best submit and use those data. Here we present recent endeavors, answer some frequently asked questions, and offer helpful tidbits to people who add data to the NDR.

INTRODUCTION

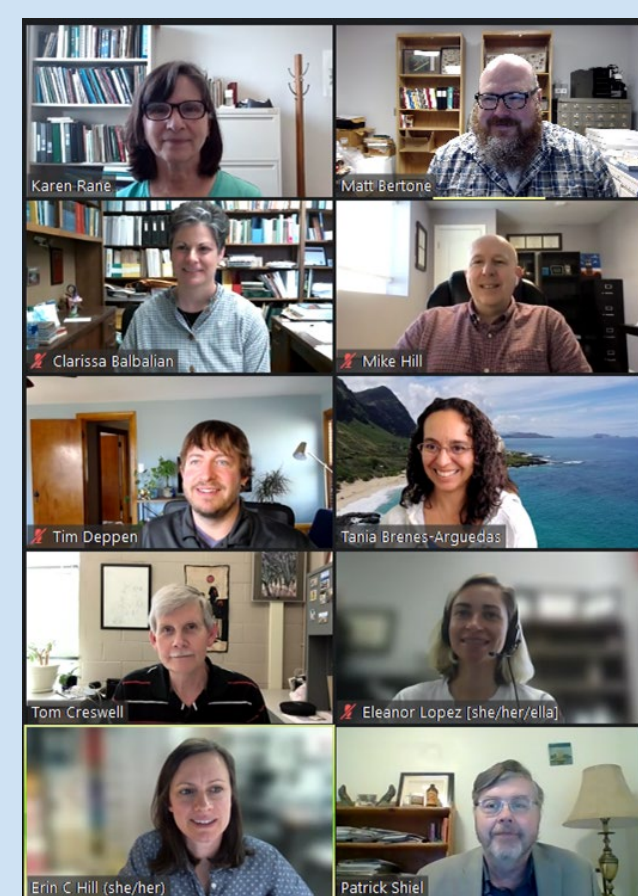
The NPDN National Data Committee functions to oversee the quality and functionality of data submitted to the National Data Repository (NDR).

Mission Statement: The mission of the National Data Committee is to support the quality and consistency of the data in the NPDN NDR by providing clear data entry guidelines, standard terminology, and guidance on problem/pest/host dictionary entries.

Goals:

1. Create data entry/upload guidelines
2. Standardize terminology used in the NDR
3. Review nomenclature for taxonomic accuracy
4. Evaluate new dictionary entry requests
5. Support development of training materials related to data in the NDR
6. Develop NPDN policy proposals as needed to support these goals

Membership: The National Data Committee comprises diagnosticians from each NPDN region, managers of the NDR, and regulatory officials (e.g. USDA-APHIS). Ideally, members represent different areas of expertise (e.g. plant pathology, entomology, weed science) in order to best address items brought to the attention of the committee.



Current members (clockwise from top left): Karen Rane, Matt Bertone, Mike Hill, Tania Brenes-Arguedas, Eleanor Lopez, Pat Shiel, Erin Hill, Tom Creswell, Tim Deppen, and Clarissa Balbalian

National Data Repository Quick Stats

First record: March 31st 2004

9 LIMS, 8 EXCEL UPLOAD LABS, 1 ONLINE DATA ENTRY LAB

80 ACTIVE DIAGNOSTIC LABS

3,908 HOST SPECIES

13,382 PEST SPECIES



Photos by Matt Bertone

2,831,786 total diagnoses!

RECENT COMMITTEE ACTIVITIES

Developing Data Upload Guidelines, including:

- "What information should be uploaded to the NDR"
- "How to Enter Diagnosis at Higher Taxonomic Levels"
- "How to Request Changes to Host/Pest Codes"
- Other topics are still under development



UPLOAD GUIDELINES

Writing articles for the NPDN Communicator

- Features on taxonomic changes and tips to choose the correct pest codes for ambiguous situations

Cleaning up "unknown" & "unidentified" codes

- Over 140 codes reviewed; removed 40+ codes that referred to 'unknowns' that were ambiguous or confusing, *with many more to go!*
- Consolidated many similar records for 'unknowns'
- Published guidelines for best practices for using unknown codes



"UNIDENTIFIED" DIAGNOSIS GUIDELINES

Reviewing requests for using data in the NDR:

- Requesters use an online form [see FAQ, right]
- Each valid request is reviewed by the data committee
- The NPDN Executive Team has final say on all requests

Are you a data geek? A taxonomy nerd?

Would you like to help us ensure NDR data is consistent and useful?

The NPDN Data Committee might be for you! We're looking for more members to be on the committee. If you're interested, please reach out to one of us at the meeting or later on.

We'd love to have your help!

FAQ

(Frequently Asked Questions)

Q. How do I upload data without a Lab Information Management System (LIMS)?

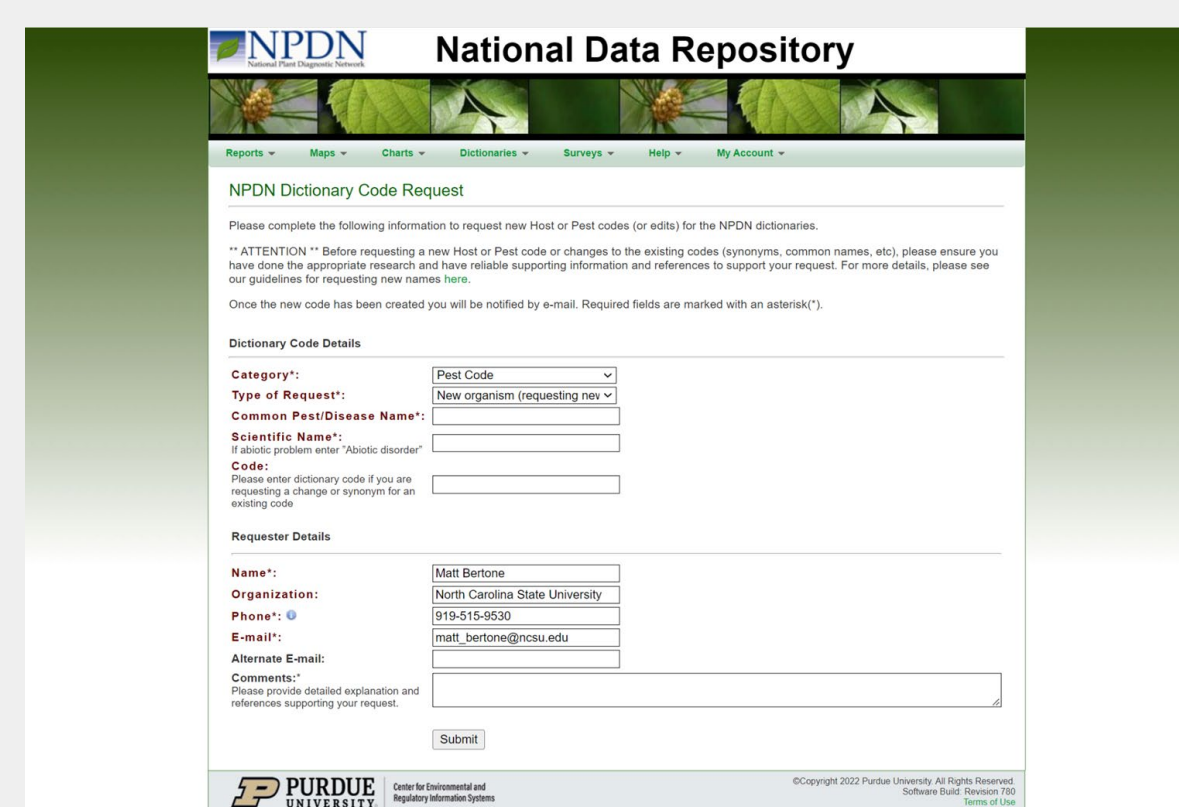
A. The best way to upload data without a LIMS is by using an Excel spreadsheet. Instructions and a template can be found using the QR link below.



DIAGNOSTIC LAB	LAB RECEIVED DATE	FIPS CODE	HOST CODE	HOST COMMON NAME	HOST
9995	2018-06-01	18157	99999	Unknown: No Site Specified	No site
9995	2018-06-01	18157	99999	Unknown: No Site Specified	No site
9995	2018-06-01	18157	99999	Unknown: No Site Specified	No site
9995	2018-06-01	18157	99999	Unknown: No Site Specified	No site
Year, Month, Day					

Q. What if I don't find the pest or host code I need?

A. New pest and host codes can be requested through most LIMS, as well as the NDR. A form can be found using the QR link below.



Q. How do I update or change pest or host codes that are incorrect or have a new taxonomy?

A. Please use the same form as above. Guidelines for how to format taxonomic names, etc. can be found by following the QR link below.



Disease Names
The NPDN/NDC uses the American Phytopathological Society's (APS) Common Names of Plant Diseases as its precedent for proposing common names to be added as part of a new pest code request.
Requests for inclusion of a new common disease name in the Pest Dictionary should conform with the name in the APS list, if available. Diseases that are not included on the APS list should conform to the following NPDN naming guidelines:
• The proposed common name should include a descriptor (usually a symptom) such as: blight, canker, dieback, rot, rust, spot, streak, wilt, etc.
• The proposed common name should not include the name of the plant, unless inclusion of a plant name is unavoidable or generally accepted (the example #1 is the only host of the disease).
• The proposed common name should not include the genus name of the pathogen.
Examples:
Incorrect: "Abate downy mildew"; Correct: "Downy mildew"
Incorrect: "Tomatoose leaf spot"; Correct: "Leaf spot"

Q. How do I get data back from the NDR?

A. All NPDN members have access to browse and download data from the NDR using their NPDN member account. Non NPDN members (or members wanting to share or publish data) can request data using the request form.



DATA REQUEST FORM



SEARCHING/NDR REPORTS VIDEO PART 1



SEARCHING/NDR REPORTS VIDEO PART 2

See side bar for more information on searching the NDR for records

Did you know?

One of the most common mistakes when entering diagnoses is searching only on the disease or pest common name and not taking time to check the pathogen or insect scientific name.

For example, did you know that in the NDR pest dictionary there are **13 different pest codes** (and *many* synonyms) for diseases having the common name of "Black Rot"?:



Black rot of grape, caused by *Phylosticta ampellicida*
Photo: Brian Olson, Bugwood.org



Black rot of apple, caused by *Botryosphaeria obtusa*
Photo: Paul Bach, Bugwood.org



Black rot of cabbage, caused by the bacterium *Xanthomonas campestris*
Photo: Gerald Holmes, Bugwood.org

It's easy to choose the incorrect entry when searching only on the common name. Which 'Black Rot' did you mean to select? And it's not just black rots, but black knots!



Photo: Purdue PPDL

Which black knot do you choose?

- Black knot
- Black knot (*Apostropha morbosus*) - pest
- Black knot (*Dibotryum morbosum*) - synonym
- Black knot (*Gibberidea sp.fpp.*) - pest



Photo: Dhandevi Pem

Furthermore, host entries can have similar errors. For example did you mean "Japanese Holly" or "Japanese Holly Fern"? Did you mean "Hemp" or "Hemp Sesbana"?

If you are having trouble understanding confidence levels and how to use them we are evaluating them this year to restrict the number of combinations and clarify their usage. *We will provide clear guidelines for users to be confident about their confidence levels!*

You can search for data in the NDR several ways using report queries:

- **Selectable Fields:** Allows you to choose what fields you want to view.
- **Summary Data:** Calculates counts of diagnoses per lab, lab methods, pest category, program and state.
- **Pest/Host Index Report:** Shows pests-host associations based on NDR data.



SELECTABLE FIELDS SEARCH



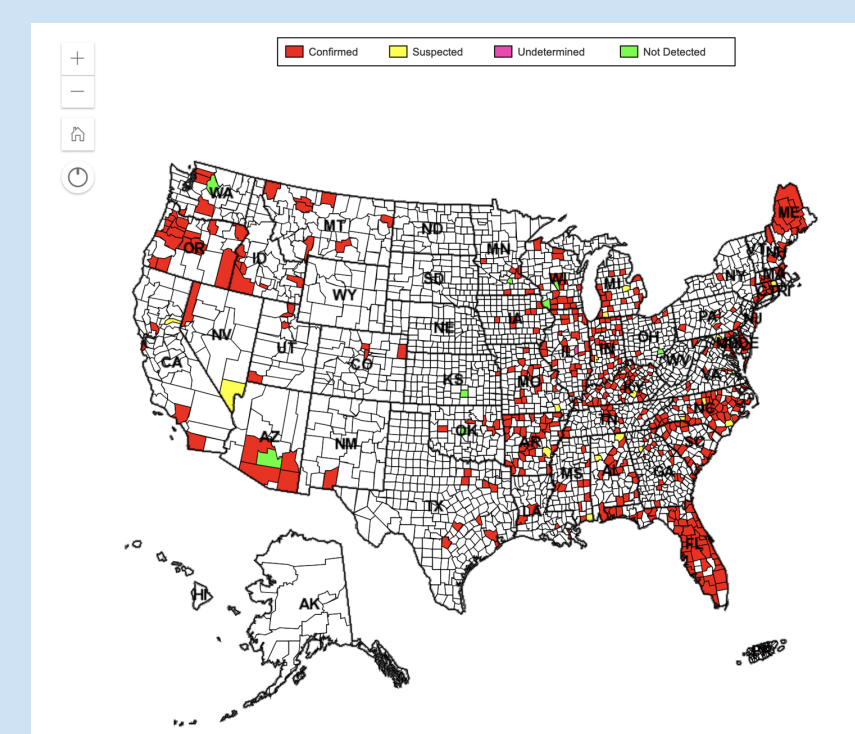
SUMMARY DATA SEARCH

Are there additional fields you want to be reported? Just let us know!

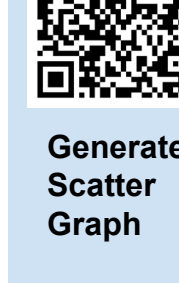


PEST/HOST INDEX REPORT

- You can even generate maps and graphs! The NDR map tool allows you to visualize where a pathogen or pest has been reported. This map shows counties with a common soil- and water-borne pathogen in 2021. **Can you guess which one?** Go to the maps tool to find out!



Generate Map



Generate Scatter Graph

Read the Fine Print: Data Use & Privacy Policy



NPDN has worked for two decades to earn the trust of our regulatory partners by ensuring confidentiality and responsible use of data.

Permission to access data in the NPDN Data Repository does not confer rights to share, present, or publish any raw, processed, or summary data from the NPDN National Data Repository (NDR). Nor does it grant rights to use data for promotional or commercial intent.

The applicant must request permission for each disclosure event whether digital or analog, oral or written from the NPDN Executive Committee who has sole authority to grant such access. Any person who wishes to share data must complete the NPDN data request form [see FAQ, left].

Users of the NDR may share data from samples in their own state and diagnosed in their lab without requesting NPDN permission. Users may also share abiotic diagnostic data without requesting permission.

The applicant is responsible for getting permission from SPHD and SPRO (of sample's state of origin) to share data from new detections and regulated organisms from the NDR.