



VERSION 10

FEB 26, 2024

OPEN  ACCESS**DOI:**

[dx.doi.org/10.17504/protocols.io.
4r3l284pql1y/v10](https://dx.doi.org/10.17504/protocols.io.4r3l284pql1y/v10)

Protocol Citation: Candace Hope Bias, Julie Haendiges, Tina Lusk Pfefer, Errol Strain, Maria Balkey, Ruth Timme 2024. NCBI submission protocol for microbial pathogen surveillance.

protocols.io

<https://dx.doi.org/10.17504/protocols.io.4r3l284pql1y/v10> Version
created by [Ruth Timme](#)

 NCBI submission protocol for microbial pathogen surveillance V.10

 In 4 collections

Candace Hope Bias¹, Julie Haendiges², Tina Lusk Pfefer³, Errol Strain³,
Maria Balkey¹, Ruth Timme²

¹Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, Maryland, USA;

²US Food and Drug Administration;

³Center for Veterinary Medicine, U.S. Food and Drug Administration, College Park, Maryland, USA

GenomeTrakr

Tech. support email: genomeTrakr@fda.hhs.gov



Ruth Timme

US Food and Drug Administration

DISCLAIMER

Please note that this protocol is public domain, which supersedes the CC-BY license default used by protocols.io.

MANUSCRIPT CITATION:

Timme, Ruth E., William J. Wolfgang, Maria Balkey, Sai Laxmi Gubbala Venkata, Robyn Randolph, Marc Allard, and Errol Strain. "Optimizing Open Data to Support One Health: Best Practices to Ensure Interoperability of Genomic Data from Bacterial Pathogens." One Health Outlook 2, no. 1 (October 19, 2020): 20. <https://doi.org/10.1186/s42522-020-00026-3>.

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working

Created: Jan 09, 2024

Last Modified: Feb 26, 2024

PROTOCOL integer ID: 93195

Keywords: NCBI submission, WGS, pathogen surveillance, GenomeTrakr

ABSTRACT

PURPOSE: This document provides detailed instructions on how to submit whole genome sequence (WGS) data and associated contextual data of microbial pathogens to NCBI for analysis within NCBI Pathogen Detection, while adhering to the INSDC standard data structure, "[Pathogen DOM](#)". The protocol includes essential steps to create a new NCBI submission environment for your laboratory group, which is crucial to have in place before data are submitted. After these initial setups, the the remaining protocol focuses on step-by-step instructions for data submission.

SCOPE: This protocol is intended for any laboratory submitting WGS data of bacterial pathogens to NCBI for analysis within NCBI Pathogen Detection. This encompasses US labs connected to GenomeTrakr, NARMS, Vet-LIRN, NAHLN, and other international networks and submitters.

GUIDANCE FOR NEW SUBMITTERS: Before initiating your first data submission, there is significant preparatory work required. We advise designating a team member to spend several days setting up the necessary systems well before your anticipated first submission.

Watch NCBI's [10min video tutorial](#) describing general submission to SRA.

ADVICE FOR FREQUENT/LARGE VOLUME SUBMISSIONS: Start by following Step 1 to establish your NCBI submission environment. For ongoing or large-scale submissions, email gb-admin@ncbi.nlm.nih.gov to arrange an account for API-based submissions.

Version history:

V9: Updated external protocol links and modified Step 2.5 to reflect change in metadata package guidance.

V8: Corrected the multi-species umbrella project accession in Step 1.5

V7: Updated guidance for creating new BioProjects, including projects for non-targeted species

V6: minor edits including updating links out to other protocols.

V5: Linking directly to the metadata template guidance instead of including duplicate copies of the files in this protocol. Updated screenshot for choosing the pathogen template to reflect changes at NCBI.

V4: updated screenshots to reflect NCBI submission portal changes. Updated custom BioSample template.

BEFORE START INSTRUCTIONS

This protocol has three sections:

- **Section 1:** Setting up NCBI accounts (for new users)
- **Section 2:** Data submission to BioSample for sample metadata and to SRA for raw reads and associated sequence metadata.
- **Section 3:** Detailed steps for creating a BioProject (usually done once during the account set-up)

Establish submission environment at NCBI

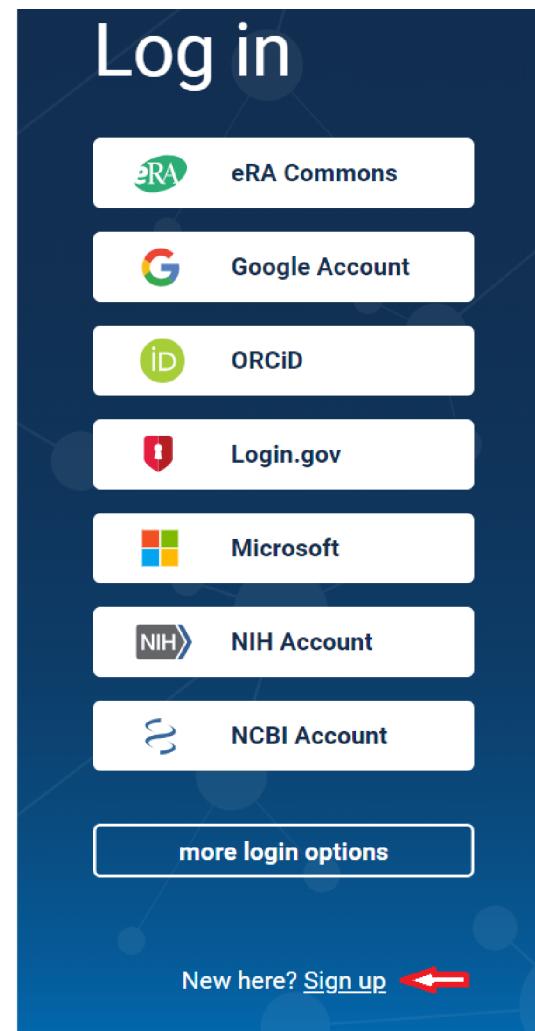
1 Set up a new NCBI submission environment for your lab:

- 1.1: Create an NCBI user account
- 1.2: Set up an NCBI submission user group for your lab
- 1.4: Bookmark the link to your submission portal
- 1.5. Identify or establish new BioProjects (detailed in **Step 3**)

Ready for data submission:

After these steps are complete you can proceed with data submission in **Step 2**.

- 1.1 **Create an NCBI user account** at NCBI: <https://www.ncbi.nlm.nih.gov/account>. This will be your own individual user account at NCBI.



The signup link is at the bottom of the page.

Choose a signup option that works for your institution.

1.2 Establish an NCBI submission user group for your laboratory.

We recommend using this user group for *all* NCBI submissions related to your lab's pathogen genome surveillance submissions.

This approach will link data submitted by your lab to the **user group** and not to individuals doing the submissions, allowing anyone in the current submission group to perform updates or retractions and answer inquiries from the NCBI staff, even if there's been a complete turnover of staff since the original data were submitted.

User groups also ensure consistent data ownership across BioProjects, BioSamples, and sequence data. If your laboratory has non-overlapping research groups submitting and

managing data at NCBI, multiple user groups can be established, if needed, to manage these efforts separately.

Your laboratory might already have a submission group established! Sign into your personal NCBI account, then check the "Group" tab in the Submission Portal ([Submission Portal | NCBI | NLM | NIH](#)). Ask your colleagues to do the same to ensure your laboratory does not already have one in place.

Group Id	Full name	Aliases	Department	Institution	Contact email	Members
fda	FDA Center for Food Safety and Applied Nutrition	FDA/CFSAN	CFSAN-ORS-DM-MMSE	US Food and Drug Administration	GenomeTrakr@fda.hhs.gov	14

View of the "Groups" tab, when selected from the NCBI Submission Portal

Click on this link to verify your membership in NCBI user groups:

<https://submit.ncbi.nlm.nih.gov/groups/>

Creating a new submission group:

1. On your NCBI profile page (<https://submit.ncbi.nlm.nih.gov/accounts/profile/>), scroll to the bottom of the page and click on the "**Create group for shared submissions**" button.

Note

The "Create group for shared submissions" button will not exist, if the user has not filled in all of the required profile information, marked with an asterisk (*) on the profile page.

Submission Portal

Personal Information

★ First (given) name	Middle name	★ Last (family) name		
Candace	Hope	Bias		
★ Email (primary)		Change primary email		
Candace.Bias@fda.hhs.gov		<input checked="" type="checkbox"/> Send email notifications on submission status change		
Email (secondary)				
★ Institution	Institution URL	★ Department		
US Food & Drug Administration		Center for Food Safety and Nutrition		
Phone 	Fax 			
★ Street	★ City	★ State/Province	★ Postal code	★ Country
Campus Drive	College Park	MD	20740	USA

[Save](#)[Back](#)

Groups

FDA Center for Food Safety and Applied Nutrition (admin, read, modify, submit, delete)

14 members: Anjanette Johnston, Yan Luo, Errol Strain, Ruth Timme, Justin Payne, Narjol Gonzalez-Escalona, Hugh Rand, Maria Balkey, Maria Hoffmann, Jayanthi Gangireddla, Julie Haendiges, Yu Wang, Elizabeth Hunter, you

[Create group for shared submissions](#)

2. On the resulting page, fill in the required information to create, at minimum, a short name, full name, and contact information for this submission group.

 Open [Invites](#) tab to invite collaborators to join your group.

* Short name	* Full name
candace.bias@fda.hhs.gov	Candace Bias's shared submissions
Aliases	Description
<input type="text"/>	<input type="text"/>
URL	
<input type="text"/>	

* Email	Alternative email			
Candace.Bias@fda.hhs.gov	<input type="text"/>			
* Institution	* Department			
US Food & Drug Administration	Center for Food Safety and Nutrition			
Phone 	Fax 			
<input type="text"/>	<input type="text"/>			
* Street	* City	* State/Province	* Postal code	* Country
Campus Drive	College Park	MD	20740	USA

Save

3. To invite members, use either the "**Invite members**" button at the top of the next, "Members" tab or from the "Invites" tab itself to navigate to the invite tab and add the invitees' emails to the text box. Click the "**Invite Members**" button when finished.

Submission Portal

Groups > Candace Bias's shared submissions

Profile Members Invites Submissions

All invited members will be granted with
read, write, submit and delete permissions,
and they will have email notifications enabled.

No pending invites.

* Emails ?

1000 characters allowed

Invite members

1.3 Managing your NCBI submission user group.

After a user group has been established it can be edited for membership and permissions by clicking in the "group" tab of the submission portal (<https://submit.ncbi.nlm.nih.gov/groups/>), then on the Group Id hyperlink, e.g "fda_ny" in the above example.

Users with admin privileges can update contact information in the "Profile" tab and membership in the "Members" tab. New members can be invited by clicking on the "Invite members" link.

The screenshot shows the 'Groups > fda' page in the National Library of Medicine Submission Portal. At the top, there's a navigation bar with links for Home, My submissions, Manage data, Groups (which is underlined), Templates, and My profile. Below the navigation is a sub-navigation bar with Profile, Members, Invites, and Submissions. A button labeled 'Invite members' is visible. To the right is a search bar with a 'Search' button. A prominent blue button at the top left of the main content area says 'Save permissions changes'. Below it is a table with columns: First name, Last name, Email, Receives email notifications, Read, Modify, Submit, Delete, Admin, and Remove member. Two rows of data are shown, each with a different set of permissions checked.

First name	Last name	Email	Receives email notifications	Read	Modify	Submit	Delete	Admin	Remove member
[Redacted]	[Redacted]	[Redacted]	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<button>Remove</button>
[Redacted]	[Redacted]	[Redacted]	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<button>Remove</button>

This user list should be kept current as members/staff enter and leave the laboratory.

Permissions levels:

- READ: primarily for collaborators who would like to view the submissions, but not edit them.
- MODIFY, SUBMIT, DELETE: Permissions to submit, modify, or retract data (members usually have all or none of these permissions)
- ADMIN: Can invite or remove members of the submission group. Ensure that at least one (or more) members of your group have ADMIN privileges.

The "Submissions" tab will show a breakdown of how many submissions have been made by this group:

All submissions	109148
api	108144
bioproject	150
biosample	573
genbank	6
sra	96
wgs	105
wgs1	8
wgs_batch	66

- 1.4** **Bookmark “My submissions” at NCBI:** <https://submit.ncbi.nlm.nih.gov/subs/>. This is the page where you view and track current and past submissions.

Submission	Title	App	Group	Status	Updated
SUB13884823	GenomeTrakr Project: University of California, Davis School of Veterinary Medicine	BioProject	fda	✓ BioProject: Processed PRJNA1024537 : GenomeTrakr Project: University of California, Davis School of Veterinary Medicine	05:00
SUB480036	Salmonella enterica Genome sequencing	BioProject	fda	✓ BioProject: Processed PRJNA1024537 : Salmonella enterica Genome sequencing	Jan 15

- 1.5** **Identify or establish new BioProjects (Umbrella and/or Data BioProjects)**

Umbrella BioProjects. If you are already part of a surveillance network, (e.g. GenomeTrakr, NARMS, Vet-LIRN, NAHLN, PulseNet, or others), you should follow the guidance from each

network coordinator for creating new BioProjects.

- For *GenomeTrakr*: create a new data project and link to one of the established umbrella BioProjects. For reference, here is a list of the major Umbrella projects for *GenomeTrakr*, organized by taxonomic classification, [PRJNA593772](#). For species not included in this list, create a general, non-targeted data BioProject and link it to PRJNA706684, our multi-species BioProject.

GenomeTrakr Umbrella projects ():

- *Campylobacter* sp. PRJNA258021
- *Cronobacter* sp. PRJNA258402
- *Escherichia coli / Shigella* PRJNA230919
- *Listeria* sp. PRJNA514048
- *Salmonella* sp. PRJNA183844
- *Vibrio* sp. PRJNA245885
- All other species PRJNA706684

If you need to establish a new umbrella BioProject, follow instructions in **Step 3** with modifications for creating a new Umbrella BioProject (**Step 3.12**).

Data BioProjects. Does your laboratory have an established data BioProject for this effort? If not, please follow the instructions in **Step 3** for creating a new one.

Note

More information:

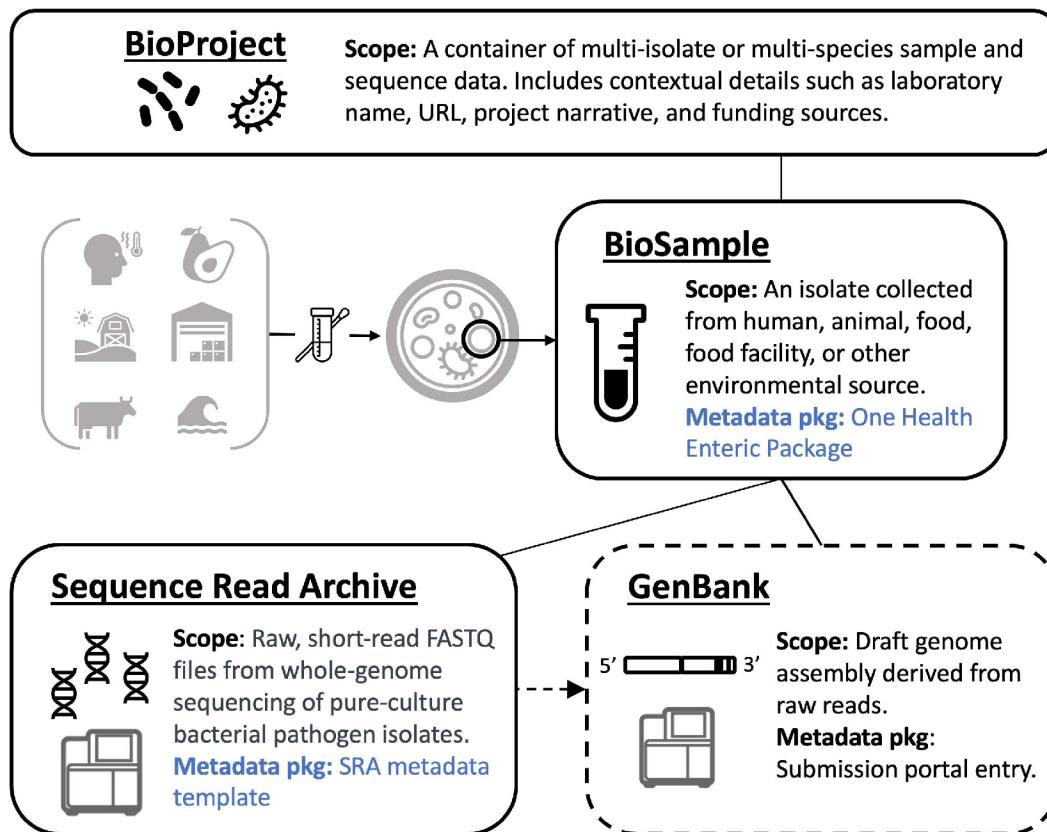
Learn more about **data vs umbrella** BioProjects in **Step 3**.

Data submission (BioSample and SRA)

2

Data submission (Sample metadata, SRA metadata, and raw sequence data), compliant with the [Pathogen DOM](#) data structure.





Overview of the Pathogen Data Object Model (Pathogen DOM) detailing the scope of BioProjects, BioSamples, and Sequence Read Archive (SRA) submissions along with associated metadata standards for enteric pathogen surveillance. Abbreviations: Pkg – Package.

Note

Arrange your submissions according to their corresponding BioProjects, ensuring that each submission workflow is dedicated to a single BioProject. In cases where your data encompass multiple BioProjects, initiate a distinct submission for each BioProject separately.

2.1 **Pre-Submission Data Quality Control**



Verify that your sequence data meet the established quality control (QC) thresholds specific to your surveillance network. Follow this protocol if this check has not yet been completed:

Protocol



NAME

Quality control assessment for microbial genomes: GalaxyTrakr MicroRunQC workflow

CREATED BY

Ruth Timme

[PREVIEW](#)

2.2

****Pre-Submission Metadata Template Validation****

Validate your sample and SRA metadata prior to uploading to NCBI using the following protocol:



Protocol



NAME

Guidance for populating and validating GenomeTrakr metadata templates (BioSample and SRA)

CREATED BY

Ruth Timme

[PREVIEW](#)

2.3

Navigate to the **My Submissions** page in the NCBI Submission Portal:<https://submit.ncbi.nlm.nih.gov/subs/>

Click "Sequence Read Archive" to start a submission.

Submission Portal

Your submissions

Start a new submission

- GenBank
- Sequence Read Archive
- Genome
- TSA
- BioProject
- BioSample
- Supplementary Files
- API

Filter / Search

From date: YYYY-MM-DD To date: YYYY-MM-DD Status: Not deleted Sort by: desc

Apps + Data archives +

Query:

110,619 submissions			
Submission	Title	App	Group
SUB14137051	CDC HAI-Seq Gram-negative bacteria, Jan 05 '24	Sequence Read Archive (SRA)	scdhec_pngt
SUB14140138	CDC NWSS - SARS-CoV-2 Wastewater Sequencing, Jan 08 '24	Sequence Read Archive (SRA)	wa_gt
SUB480036	Salmonella enterica Genome sequencing	BioProject	fda

1 2 3 4 5 6 7 8 9 ... 4,422 4,423 4,424 4,425 Next >

2.4 Click the “New submission” button.

Submission Portal

Sequence Read Archive (SRA)

New submission 

Short description and brief instructions

Options to upload data:

- Upload via Cloud: Amazon S3 or Google Cloud
- Upload via Aspera command line or FTP UPDATED OCTOBER 2023
- Upload via Web browser or Aspera browser plugin UPDATED OCTOBER 2023

Filter / Search

From date: YYYY-MM-DD To date: YYYY-MM-DD Status: Not deleted Sort by: desc

Data archives +

Query:

1 2 3 4 5 6 7 8 9 ... 38 39 40 41 Next >

2.5 SUBMITTER tab:

Populate with submitter info. The “submitter” is the name of the person AND user group, who is physically doing the submissions, not a supervisor or PI.

Select the appropriate submission group name (see **Step 1.2** for creating a new submission group), and describe the submitting organization or laboratory name. This will be auto-populated from the contact info you included in your NCBI user account. Click "Continue" to proceed.

2.6 GENERAL INFO tab:

1. BioProject: Did you already have a data BioProject for this effort? If not please follow instructions in **Step 3** for creating a new data or umbrella BioProject. Return back to this sub-step with the **data BioProject** accession in hand.

Click "Yes" and paste in your data BioProject accession, e.g. PRJNA614995. Note: Be sure not to use an umbrella BioProject. Select the appropriate BioProject under the umbrella. Otherwise, you will receive an error and not be able to proceed.

2. BioSample: Click "NO" here. You will be registering BioSamples within this current submission.

3. Release date: Choose "Release immediately following processing".

4. Click Continue.

Sequence Read Archive (SRA) submission: SUB14157251

Listeria monocytogenes Raw sequence reads, Jan 17 '24

1 SUBMITTER 2 GENERAL INFO 3 BIOSAMPLE TYPE 4 BIOSAMPLE ATTRIBUTES 5 SRA METADATA 6 FILES 7 REVIEW & SUBMIT

General Information

BioProject

BioProject describes the goal of your research effort.

* Do you already have a BioProject accession number for this research?

Yes No (a BioProject will be created within this submission)

* Existing BioProject

PRJNA295367

PRJNA295367: Listeria monocytogenes Raw sequence reads

BioSample

The BioSample records the detailed biological and physical properties of the sample that was sequenced. A BioSample can be used in more than one BioProject since it should be used for all the data that were obtained from that sample. Usually SRA data sets are generated from more than one sample.

* Do you already have BioSample accession numbers for these samples?

Yes No (BioSamples will be created within this submission)

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

Release immediately following processing
 Release on specified date or upon publication, whichever is first

Please allow 24-48 hours for propagation of the data to the NCBI SRA public site.

Continue

Example of filled in "General Info" tab. Please use the BioProject accession necessary for your organism and project.

2.7 BIOSAMPLE TYPE tab:

Choose the appropriate metadata package here for your sample (which sample template did you populate?)

For GenomeTrkr, we recommend using the **One Health Enteric** package (custom templates for US public health submitters available [here](#)).

Submission Portal

Sequence Read Archive (SRA) submission: SUB14157251

Listeria monocytogenes Raw sequence reads, Jan 17 '24

1 SUBMITTER > 2 GENERAL INFO > 3 BIOSAMPLE TYPE > 4 BIOSAMPLE ATTRIBUTES > 5 SRA METADATA > 6 FILES > 7 REVIEW & SUBMIT

Sample Type

* Select the package that best describes your samples.

All packages Packages for MAG submitters Packages for metagenome submitters

(Optional) Filter packages by organism name

Enter the full scientific name of your samples, e.g., Escherichia coli

[Reset and show all packages](#)

- ① To filter for relevant BioSample packages, enter the full scientific name of the organism of your samples.
- If your BioSamples are derived from a species not represented in NCBI's Taxonomy database, enter the genus-level name, e.g., *Escherichia*
 - If your BioSamples are derived from more than one organism, enter the common species, genus, or family, e.g., *Enterobacteriaceae*
 - If your BioSamples are metagenomic/environmental, or metagenome-assembled genomes (MAG), select the appropriate tab above
 - For more information about organism names, see [Organism information](#).

NCBI packages [More...](#)

- SARS-CoV-2: clinical or host-associated**
Use for SARS-CoV-2 samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- SARS-CoV-2: wastewater surveillance**
Use for SARS-CoV-2 wastewater surveillance samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of SARS-CoV-2 cases.
- Pathogen**
Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.
- One Health Enteric**
Use for microbial isolates that are collected for genomic surveillance of enteric pathogens. Sample spaces include the following: 1. human/animal hosts; 2. food samples; 3. food facilities; 4. environmental samples (farm, water, and the environment). US public health agencies have created customized versions of this package that include more specific guidance, controlled vocabulary picklists, and sub-packages for each of the 4 sample types.
 - [GitHub repository](#)
 - [Validation for the OHE package](#)

GSC MiS packages for genomes, metagenomes, and marker sequences [More...](#)

- MIGS Cultured Bacterial/Archaeal**
Use for cultured bacterial or archaeal genomic sequences. Organism must have lineage [Bacteria](#) or [Archaea](#).
- MIGS Eukaryotic**
Use for eukaryotic genomic sequences. Organism must have lineage [Eukaryota](#).
- MIGS Viral**
Use for virus genomic sequences. Organism must have lineage [Viruses](#).
- MIMAG Metagenome-assembled Genome**
Use for metagenome-assembled genome sequences produced using computational binning tools that group sequences into individual organism genome assemblies starting from metagenomic data sets. Organism cannot contain the term 'metagenome'. Use the MIUVIG package for virus genomes. Before creating BioSamples for prokaryotic and eukaryotic MAGs, please read and follow the [MAG submission instructions](#).
- MIMARKS Specimen**
Use for any type of marker gene sequences, e.g. 16S, 18S, 23S, 28S rRNA or COI obtained from cultured or voucher-identifiable specimens. Organism cannot contain the term 'metagenome'.
- MIMARKS Survey related**
Use for any type of marker gene sequences, e.g. 16S, 18S, 23S, 28S rRNA or COI obtained directly from the environment, without culturing or identification of the organisms. Organism must be a

Example "BioSample Type" tab. Note "One Health Enteric," recommended for GenomeTrakr submissions, has been selected.

2.8 BIOSAMPLE ATTRIBUTES tab:

Choose "Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples".

Then click "Choose File" and browse to your populated metadata template.

Note

If you have not yet populated and validated your GenomeTrakr BioSample metadata template,  [go to step #2](#)

Antibiogram data: please provide if you have it!

Click "**Continue**".

Sequence Read Archive (SRA) submission: SUB14157251

Listeria monocytogenes Raw sequence reads, Jan 17 '24

1 SUBMITTER 2 GENERAL INFO 3 BIOSAMPLE TYPE 4 BIOSAMPLE ATTRIBUTES 5 SRA METADATA 6 FILES 7 REVIEW & SUBMIT

Attributes

Package One Health Enteric; version 1.0

* How do you want to provide your BioSample attributes?

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

* Attributes file

OneHealthEntericMetadata_listeria_240118.xlsx 144.4 kB 2024-01-18 10:44

Delete

- Template for BioSample package One Health Enteric; version 1.0

[Download Excel](#) or [download TSV](#)

For column explanations and examples, please see the [sample attributes page](#).

For more information, please see [creating sample attribute file](#).

Antibiogram – pathogen MIC (optional)

[Choose file](#) or drag and drop it here

- [Download Antibiogram Excel template](#)

Once the spreadsheet is completed, save as a 'Text (Tab-delimited)' file, and upload it.

For more information, please see [providing pathogen MIC antibiogram data](#).

Antibiogram – mycobacterial non-MIC (optional)

[Choose file](#) or drag and drop it here

- [Download Antibiogram Excel template](#)

Once the spreadsheet is completed, save as a 'Text (Tab-delimited)' file, and upload it.

For more information, please see [providing mycobacterial non-MIC antibiogram data](#).

[Continue](#)

NCBI will do a validation check on your metadata. Resolve any red "errors" reported back by editing the spreadsheet and replacing the uploaded file. Review any yellow "Warnings" and proceed if everything looks ok.

Note

If you are using the One Health Enteric Package BioSample metadata template downloaded from the CFSAN Biostatistics GitHub and receiving an error like the one below, an the empty, original version of the template may have been uploaded. Try again with your completed template.

Sequence Read Archive (SRA) submission: SUB14159955

PulseNet Salmonella enterica Genome sequencing, Jan 18 '24



Attributes



Error: Non-empty value in column without header: "Version Number: 1.3"

Note

If you have followed One Health submission guidance and included the sub-species and serovar to the "Organism name" field, you may see the warning pictured below. You do not need to do anything in response to this warning. It exists merely to tell you that the "sub species" and "serovar" fields have been created in addition to the "Organism name" field. The metadata will be preserved in the Organism name as well as used to populate the new serovar and sub species attributes.



Warning: Provided taxonomy information was revised according to NCBI Taxonomy database rules. Please contact biosamplehelp@ncbi.nlm.nih.gov if you have any questions.

Sample name	Organism name	Changed	Attribute name	Attribute value
SAL-23-VL-LA-PA-RYAN-0002	Salmonella enterica subsp. enterica serovar 4,[5],12:i:-	Yes	serovar	4,[5],12:i:-
SAL-23-VL-LA-PA-RYAN-0002	Salmonella enterica subsp. enterica serovar 4,[5],12:i:-	Yes	sub species	enterica
SAL-23-VL-LA-PA-RYAN-0003	Salmonella enterica subsp. enterica serovar Dublin	Yes	serovar	Dublin
SAL-23-VL-LA-PA-RYAN-0003	Salmonella enterica subsp. enterica serovar Dublin	Yes	sub species	enterica

Click "Continue".

2.9 SRA METADATA tab:

Choose: "Upload a file using Excel or text format (tab-delimited)".

Sequence Read Archive (SRA) submission: SUB9477908

GenomeTrakr Project: US Food and Drug Administration, Apr 14 '21

The screenshot shows the 'SRA metadata' step of the SRA Submission Wizard. At the top, there is a navigation bar with steps 1 through 7: 1 SUBMITTER, 2 GENERAL INFO, 3 BIOSAMPLE TYPE, 4 BIOSAMPLE ATTRIBUTES, 5 SRA METADATA (which is highlighted in blue), 6 FILES, and 7 REVIEW & SUBMIT. Below the navigation bar, the title 'SRA metadata' is displayed. A note says: 'For more detailed help with SRA submission please read the [SRA Submission Wizard Help](#)'. A section titled 'How do you want to provide your metadata?' contains two options: 'Use built-in table editor' (radio button not selected) and 'Upload a file using Excel or text format (tab-delimited)' (radio button selected). Below this, a table lists a single metadata file: 'SRA_metadata-GTexample.xlsx' (53.7 kB, 2021-04-14 17:11). There is a 'Delete' button next to the file name. A note below the table says: 'Use one of these ways to fill out SRA metadata: Download Excel spreadsheet (designed to make it easier to select the correct metadata values), edit, save and then upload the modified Excel file.' At the bottom of the page is a large blue 'Continue' button.

Upload your populated **SRA metadata template** ([go to step #2](#) for where to get this file)

Click "Continue".

NCBI will do a validation check on your sequence metadata. Resolve any red "errors" reported back by editing the spreadsheet and replacing the uploaded file. Review any yellow "Warnings" and proceed if everything looks ok.

Click "Continue".

2.10 Files tab:

Each laboratory will establish its own path for transferring files. Select the radio button corresponding to the means you will use.

In general, selecting the web browser option should work for uploading ~48 sequences at a time. For a more stable internet connection, your laboratory can use FTP or Aspera. Directions for doing so pop up after clicking the FTP radio button. Firewalls may prevent use of Aspera or AWS routes of submission.

Note

It is generally not recommended to check the "Autofinish submission" box as this would not allow you to make corrections, if needed.

Sequence Read Archive (SRA) submission: SUB14157251

Listeria monocytogenes Raw sequence reads, Jan 17 '24



Files

- ⓘ • Each file must be listed in the [SRA metadata table you uploaded](#). If you are uploading a **tar** archive, list each file name, not the archive name.
- ⓘ • Unique file names that do not contain any sensitive information should be used for all files. File names as submitted appear publicly when data is retrieved from the cloud.
- ⓘ • Files can be compressed using **gzip** or **bzip2**, and may be submitted in a **tar** archive, but archiving or compressing your files is not required. **Do not use zip!**

★ How do you want to provide files for this submission?

- Web browser upload via HTTP or Aspera Connect plugin
Do not use web browser HTTP upload if you are uploading files over 10 GB or more than 300 files.
- FTP or Aspera Command Line file preload
All files for a submission must be uploaded into a single folder.
- AWS or GCP bucket

Autofinish submission

- ⓘ Try to complete the submission automatically. Basic checks will be done on the files in the background. If there are no errors, this submission will be submitted for you. If the checks fail, you will get email notification to come back to fix the problems.

Continue

2.11 REVIEW & SUBMIT tab:

Check over your entire submission, then click "**Submit**."

If corrections are needed, you can go back and select individual tabs to edit your submission.

Note

If you are having trouble finalizing your submission, contact the relevant NCBI database for assistance, and include your submission ID in the email subject (SUB#####):

BioSample (for source metadata issues): biosamplehelp@ncbi.nlm.nih.gov
SRA (for raw sequence or sequence metadata issues): sra@ncbi.nlm.nih.gov

2.12 BioSample accessions:

BioSample accessions will be automatically created upon submission and will be available on the "my submissions" page of the Submission portal by clicking on "## objects" within the submission record. You can also download by clicking the "Download attributes file with BioSample accessions". Accessions will start with SAMNxxxxxxxx. You will also receive an email within 12 hours, but typically much faster, containing these same accessions.

2.13 SRA Accessions:

SRA run accessions will be available on the “My Submissions” page of the Submission portal by clicking on “## objects” within the submission record. You can also download by clicking the “Download metadata file with SRA accession”. Accessions will start with SRRxxxxxx.” You will also receive an email with these same accessions within 24 hours, but typically much faster, containing these same accessions.

30 submissions				
Submission	Title	Group	Status	Updated
SUB5616822	GenomeTrakr Project: US Food and Drug Administration, May 14 '19	fda	✓ SRA: Processed (2 objects) • SRR9052981 • SRR9052982 Download metadata file with SRA accessions View and manage my SRA submission data	May 14
SUB5110190	GenomeTrakr Project: Bacillus cereus, New York State Department of Health, Wadsworth Center, Jan 31 '19	fda_ny	✓ SRA: Processed (6 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 31
SUB5106005	GenomeTrakr Project: NY State Dept. of Health, Wadsworth Center, Jan 30 '19	fda_ny	✓ SRA: Processed (20 objects) Download metadata file with SRA accessions View and manage my SRA submission data	Jan 30

2.14 Important data stewardship and curation notes:

- Develop an internal method for storing and tracking your BioSample and SRR accessions! They are required for making future updates to your records.
- For updates, corrections, or retractions to your BioSample and SRA records, follow the guidance provided in the NCBI Curation Protocol. Some edits can be made within the submission portal and others need to be done via email.

Protocol



NAME

NCBI Bacterial Pathogen Data Curation Protocol: SOP for Editing GenomeTrakr Submissions

CREATED BY

Ruth Timme

[PREVIEW](#)

Safety information

Caution: It is possible for a single BioSample to have more than one SRR ID. Two scenarios include:

1. Two runs were submitted for the same isolate/BioSample, which is not generally recommended for surveillance. Follow **Step 3** in the NCBI curation protocol to retract one of them).
2. if the initial submission was retracted and a new run was submitted. It's important to keep track of both IDs, even if one was retracted.

BioProject Creation

3 Create a new BioProject

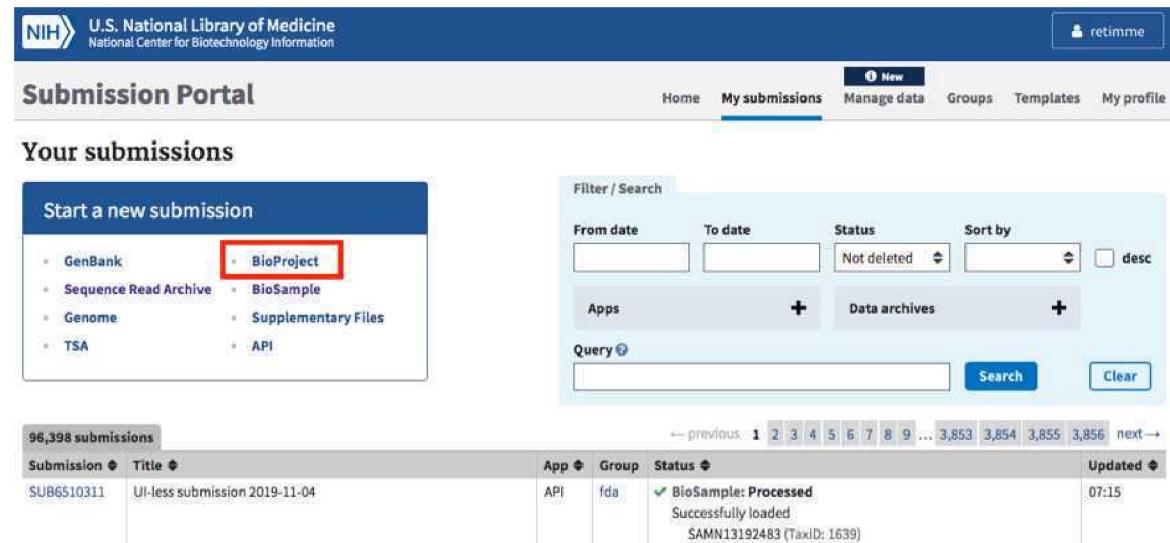
BioProjects are an organizing tool at NCBI that pulls together different kinds of data submitted across multiple NCBI databases. Each BioProject has a unique URL, providing a home page with a title, description, links to lab websites, publications, and funding resources associated with a particular project, along with links to the deposited data. A basic **data BioProject** holds actual sequence data, assemblies, and their associated metadata. An **umbrella BioProject** is a way to group two or more data BioProjects together, which is useful for coordinating disease surveillance and for looking across the grouped BioProjects in a single view.

This portion of the protocol describes the steps for creating a new data BioProject linked to an existing umbrella BioProject (usually established by a coordinating group, e.g. GenomeTrakr, NARMS, Vet-LIRN).

*If you need to create a new Umbrella BioProject, modifications are summarized in Step **3.12**.

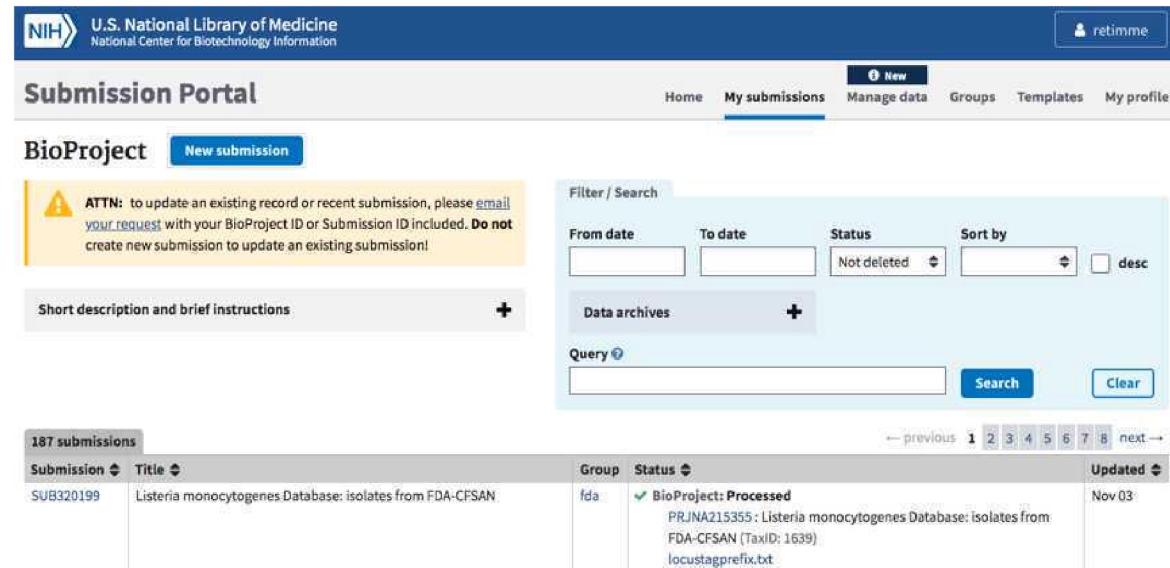
3.1

Navigate to the “My Submissions” page, <https://submit.ncbi.nlm.nih.gov/subs/>, and click “BioProject” in the “Start a new submission” box.



The screenshot shows the NCBI Submission Portal interface. At the top, there's a navigation bar with the NIH logo, "U.S. National Library of Medicine", and "National Center for Biotechnology Information". On the right side of the top bar, there are links for "Home", "My submissions" (which is underlined in blue), "Manage data", "Groups", "Templates", and "My profile". Below the top bar, the main title "Submission Portal" is displayed, followed by "Your submissions". A sidebar on the left contains a "Start a new submission" section with several options: GenBank, BioProject (highlighted with a red box), Sequence Read Archive, BioSample, Genome, Supplementary Files, TSA, and API. To the right of the sidebar is a search/filter panel titled "Filter / Search" with fields for "From date", "To date", "Status" (set to "Not deleted"), and "Sort by". Below the search panel are two categories: "Apps" and "Data archives", each with a plus sign to expand. A "Query" input field and a "Search" button are at the bottom of the filter panel. The main content area shows a table of "96,398 submissions". The columns are "Submission", "Title", "App", "Group", "Status", and "Updated". One row is visible: "SUB6510311" with title "UI-less submission 2019-11-04", App "API", Group "fda", Status "BioSamples: Processed Successfully loaded SAMN13192483 (TaxID: 1639)", and Updated "07:15".

3.2 Click the “New submission” box:



The screenshot shows the NCBI Submission Portal interface, specifically the "BioProject" section. At the top, the NIH logo, "U.S. National Library of Medicine", and "National Center for Biotechnology Information" are present, along with the "My submissions" tab which is underlined in blue. Below the top bar, the main title "Submission Portal" is displayed, followed by "BioProject". A "New submission" button is visible. A yellow warning box contains the text: "ATTN: to update an existing record or recent submission, please [email your request](#) with your BioProject ID or Submission ID included. Do not create new submission to update an existing submission!" To the right of the warning box is a search/filter panel titled "Filter / Search" with fields for "From date", "To date", "Status" (set to "Not deleted"), and "Sort by". Below the search panel are two categories: "Data archives" and "Query" input field with a "Search" button and a "Clear" button. The main content area shows a table of "187 submissions". The columns are "Submission", "Title", "Group", "Status", and "Updated". One row is visible: "SUB320199" with title "Listeria monocytogenes Database: isolates from FDA-CFSAN", Group "fda", Status "BioProject: Processed PRJNA215355: Listeria monocytogenes Database: isolates from FDA-CFSAN (TaxID: 1639) locustagprefix.txt", and Updated "Nov 03".

3.3 *Submitter tab:*

Populate with submitter info. An NCBI "submitter" is the name of the person or submission group who is managing the submissions, not a supervisor or PI.

Select the appropriate submission group name (see **Step 1.2** for creating a new submission group), and describe the submitting organization or laboratory name. This will be auto-populated from the contact info you included in your NCBI user account.

3.4 **Project type tab:**

Project data type: Genome sequencing and assembly.

Sample scope:

For a **Data BioProject**. Select multi-species. This will allow you to submit multiple different species to the BioProject.

3.5 **Target tab:**

For a **Data BioProject**. Populate ONLY the Organism name here:

For targeted-pathogen BioProjects:

Organism name = Include a Genus name, e.g., *Salmonella* sp.

For non-targeted pathogens

Organism name = "bacteria"

Create a description of the scope of the project (e.g. "enteric bacteria").

3.6 **General info tab:**

Click "Release immediately following processing".

Include a brief title describing the effort.

- **Data BioProject Title:** e.g., "GenomeTrakr Project: NY State Dept. of Health, Wadsworth Center".

Public Description: e.g., "Whole-genome sequencing of pure-cultured microbial pathogens as part of XXXX surveillance effort."

Relevance: environmental.

Is your project part of a larger initiative that is already registered at NCBI?

- **Data BioProjects.** Click "Yes" and include a brief description and umbrella BioProject accession number (see **Step 1.5**). This will properly link your data project to the umbrella.

Note

Note: We advise against linking data BioProjects to multiple umbrella BioProjects.

BioProject submission: SUB12915263

FDA BioProject: bacterial pathogens

 This submission was copied from [SUB12914688](#).

1 SUBMITTER > **2 PROJECT TYPE** > **3 TARGET** > **4 GENERAL INFO** > **5 BIOSAMPLE** > **6 PUBLICATIONS** > **7 REVIEW & SUBMIT**

General Info

Release date

 Note: Release of BioProject or BioSample is also triggered by the release of linked data.

*** When should this submission be released to the public?**

- Release immediately following processing
- Release on specified date or upon publication, whichever is first

*** Project title** 

FDA BioProject: bacterial pathogens

*** Public description** 

Whole genome sequencing of pure-cultured bacterial pathogens for FDA surveillance and outbreak investigations.

Relevance 

Environmental 

*** Is your project part of a larger initiative which is already registered with NCBI?**

- No
- Yes (not very common)

*** Initiative description**

GenomeTrakr umbrella for diverse pathogen species

*** BioProject accession**

PRJNA706684

 If you are registering a project that is part of an initiative which is already registered in the BioProject database, then please tell us the existing BioProject accession and provide a general description of the larger initiative. For example, the ENCODE project and Human Microbiome project have several subprojects. This information is needed for project linking.

External links

Description 

URL 

Delete 

 Add another link

Select your grants

 Use this tool to look up grants from many subscribed governmental funding agencies (eg NIH, CDC, FDA and VA) and some non-governmental funding sources (eg HHMI). You can search by grant number, title or grantee name. If your grant is not included, you can select the "Add grants manually" option within this tool to add your grant.

 Add grants

Consortium name 

Consortium URL 

Data provider 

Data provider URL 

Delete 



3.7 ***BioSample tab:***

Leave blank!! You will create biosamples separately.

3.8 ***Publications tab:***

If relevant, include publications from your laboratory.

3.9 ***Review and Submit tab:***

Check if everything looks correct and edit if necessary, then click "**submit**."

Submission Portal

BioProject submission: SUB12915263

FDA BioProject: bacterial pathogens



This submission was copied from SUB12914688.

1 SUBMITTER | **2 PROJECT TYPE** | **3 TARGET** | **4 GENERAL INFO** | **5 BIOSAMPLE** | **6 PUBLICATIONS** | **7 REVIEW & SUBMIT**

Review & Submit

This BioProject submission will be released **immediately following processing**.

Submitter

Submitter Ruth Timme
ruth.timme@fda.hhs.gov

Submitting organization US Food and Drug Administration

Project type

Sample scope Multispecies
Target description Bacterial pathogens not included in the major taxa scoped for GenomeTrkr.

Target

Organism name (taxid) bacterial pathogens

General information

Project details

Parent project ID PRJNA706684
Parent project description GenomeTrkr umbrella for diverse pathogen species
Project type genome sequencing
Title FDA BioProject: bacterial pathogens
Description Whole genome sequencing of pure-cultured bacterial pathogens for FDA surveillance and outbreak investigations.
Relevance Environmental

Example for a new non-targeted BioProject

3.10

The **BioProject accession** will be available within a few minutes on the “My Submissions” page of the Submission portal in the format “PRJNAxxxxxx.” You will also receive an email containing the new accession.

The screenshot shows the NIH Submission Portal interface. At the top, there's a navigation bar with the NIH logo, "U.S. National Library of Medicine", and "National Center for Biotechnology Information". The main area is titled "Submission Portal" and "Your submissions". On the left, there's a sidebar with a "Start a new submission" button and a list of submission types: GenBank, BioProject, Sequence Read Archive, BioSample, Genome, Supplementary Files, TSA, and API. The main content area displays a table of submissions. One row is highlighted with a red arrow pointing to the "Status" column, which shows "BioProject: Processed". Other columns include "App", "Group", "Status", and "Updated". The table has a total of 32,767 submissions. The bottom right corner of the screenshot shows the date "16/22".

- 3.11** If you are part of a coordinated surveillance effort, like GenomeTrakr, please alert the coordinating body that a new BioProject was created under an existing umbrella.

For GenomeTrakr, contact genometrakr@fda.hhs.gov

- 3.12 Creating a new Umbrella BioProject:**

Proceed as outlined in the above steps with the following modifications:

PROJECT TYPE tab:

For an **Umbrella BioProject**. Select multi-species. This will allow you to link multiple data BioProjects representing different species under a single umbrella.

TARGET tab:

For an **Umbrella BioProject**. Leave the Organism name field blank. Include a list or description of species you intend to include in this effort. E.g. "bacterial foodborne pathogens", or "SARS-CoV-2"

GENERAL INFO tab:

Umbrella BioProject Title: e.g. "Microbial pathogen surveillance at NY State Dept. of Health, Wadsworth Center."

Is your project part of a larger initiative that is already registered at NCBI?

- For an Umbrella BioProject: click "NO"
-

The last step is to email bioprojecthelp@ncbi.nlm.nih.gov:

Example email:

Note

"Dear BioProject and PD help teams,

Please convert the PRJNA##### to an Umbrella BioProject. Our laboratory will be submitting data under the XXX effort (SARS-CoV-2, GenomeTrakr, Vet-LIRN, NARMS, HAI, or more general pathogen surveillance).

I'd be happy to provide any additional details you might need.

Thank you,"

After the conversion is complete you can use the new Umbrella accession to properly link any new data BioProjects being created.

3.13 Important data stewardship and curation notes:

- Develop an internal method for storing and tracking your BioProject accessions! They are required for every BioSample and sequence data submission to ensure proper linkage.
- Bookmark URLs for each of your data BioProjects to monitor the public-facing view of your submissions.
e.g. Virginia DCLS's GenomeTrakr *Salmonella* BP:
<https://www.ncbi.nlm.nih.gov/bioproject/219491>
- For updates to your BioProjects, follow the guidance provided in the NCBI Curation Protocol. Some edits can be made within the submission portal and others need to be done via email.
 [go to step #2.14](#) NCBI Data Curation Protocol.