Jun 26, 2024

NCBI submission protocol for HPAI milk surveillance

Forked from NCBI submission protocol for microbial pathogen surveillance

DOI
dx.doi.org/10.17504/protocols.io.q26g711b1gwz/v1

Ruth Timme[1]

[1]US Food and Drug Administration

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

Ruth Timme
US Food and Drug Administration

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.q26g711b1gwz/v1

Protocol Citation: Ruth Timme 2024. NCBI submission protocol for HPAI milk surveillance. **protocols.io** **https://dx.doi.org/10.17504/protocols.io.q26g711b1gwz/v1**

Manuscript citation:

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: In development
We are still developing and optimizing this protocol.

Created: January 09, 2024

Last Modified: June 26, 2024

Protocol Integer ID: 100855

Keywords: NCBI submission, pathogen surveillance, H5N1, HPAI, Influenza A, milk

Disclaimer

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

Abstract

PURPOSE: This DRAFT protocol provides detailed instructions on how to submit raw targeted amplicon sequence data and associated contextual data of H5N1 to NCBI. 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

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:

Before start

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)

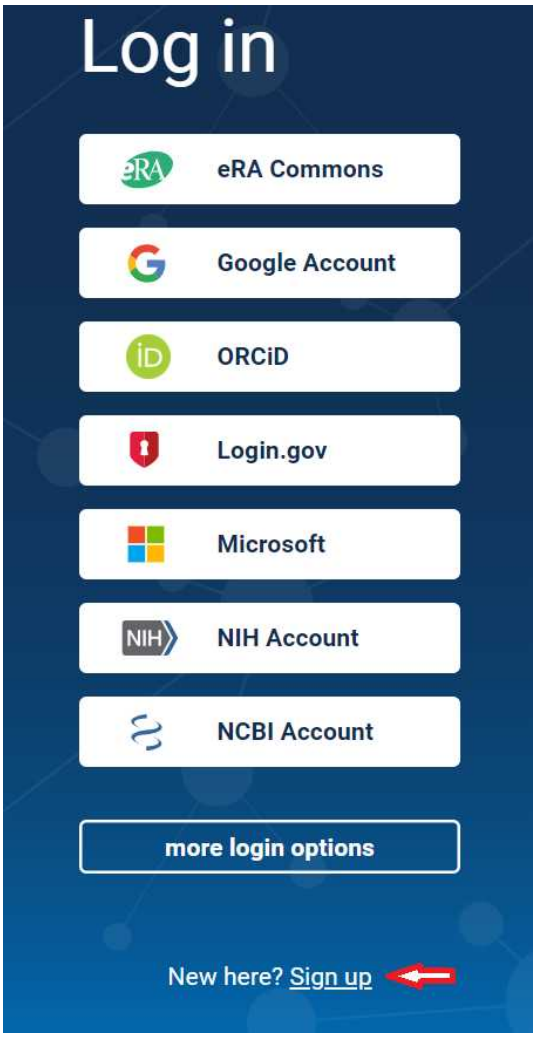
Establish submission environment at NCBI

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.**

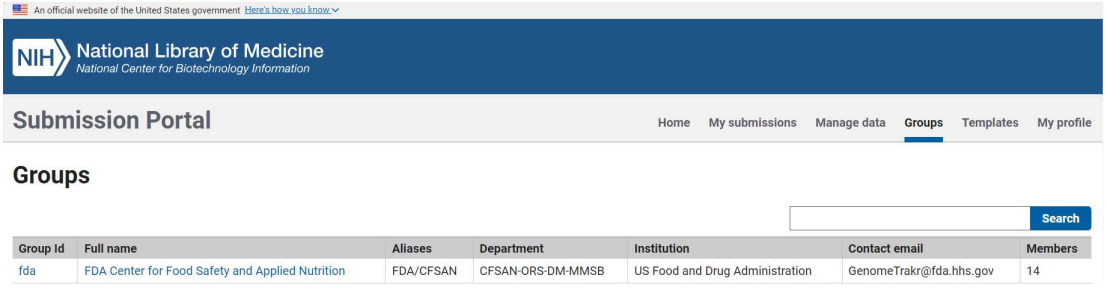
We recommend using this user group for *all* NCBI submissions related to your labs's pathogen genome surveillance submissions.

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.



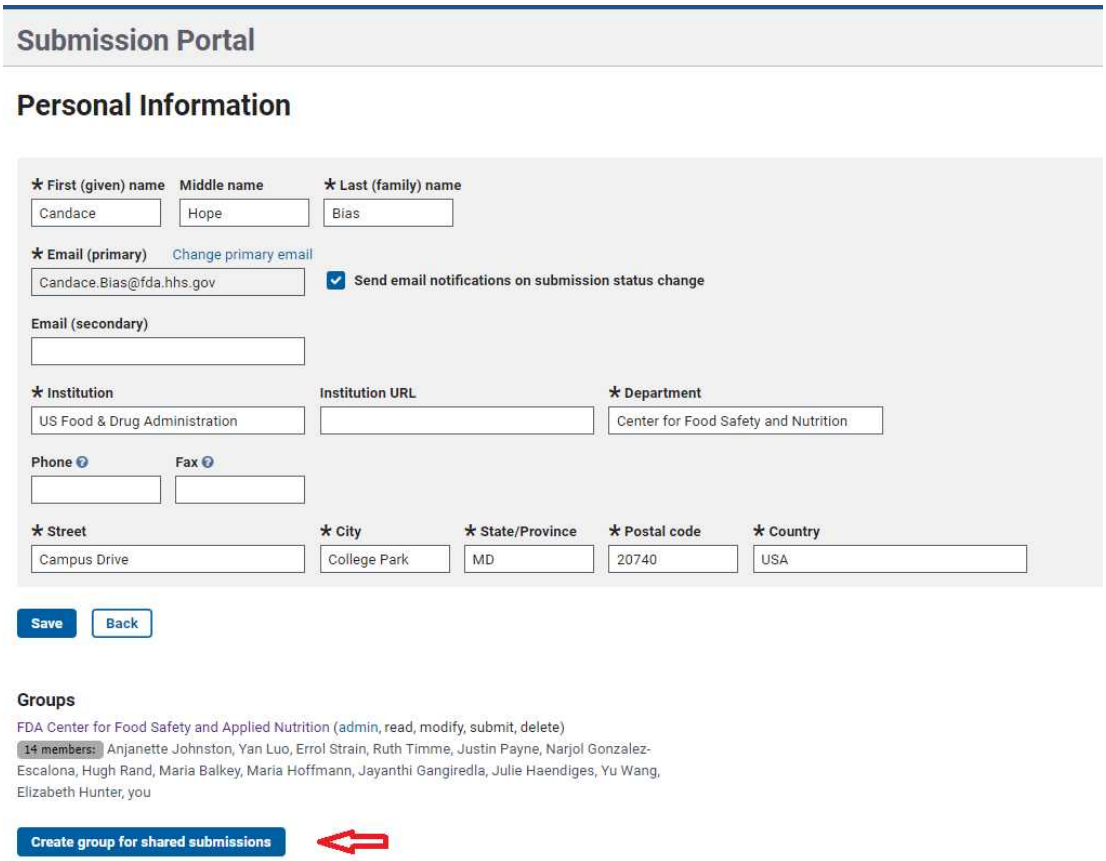
View of the "Groups" tab, when selected from the NCBI Submission Portal

Click on this link to verify your membership in NCBI user groups:
https://submit.ncbi.nlm.nih.gov/groups/

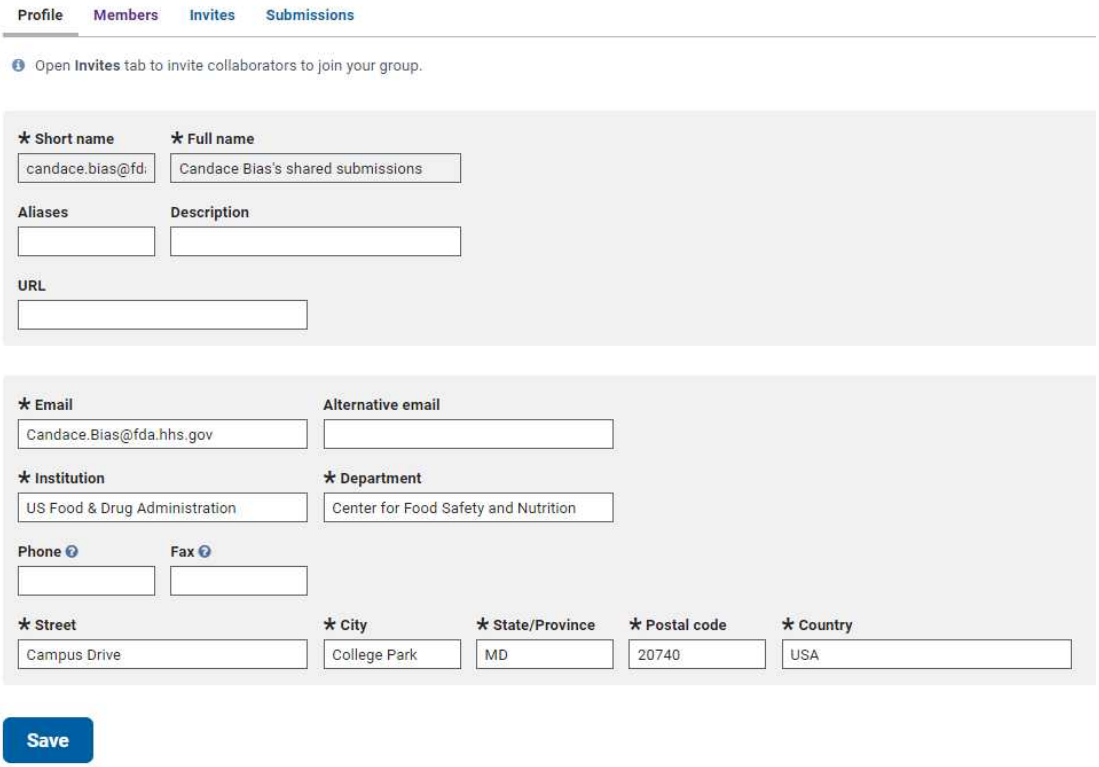
Creating a new submission group:

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.

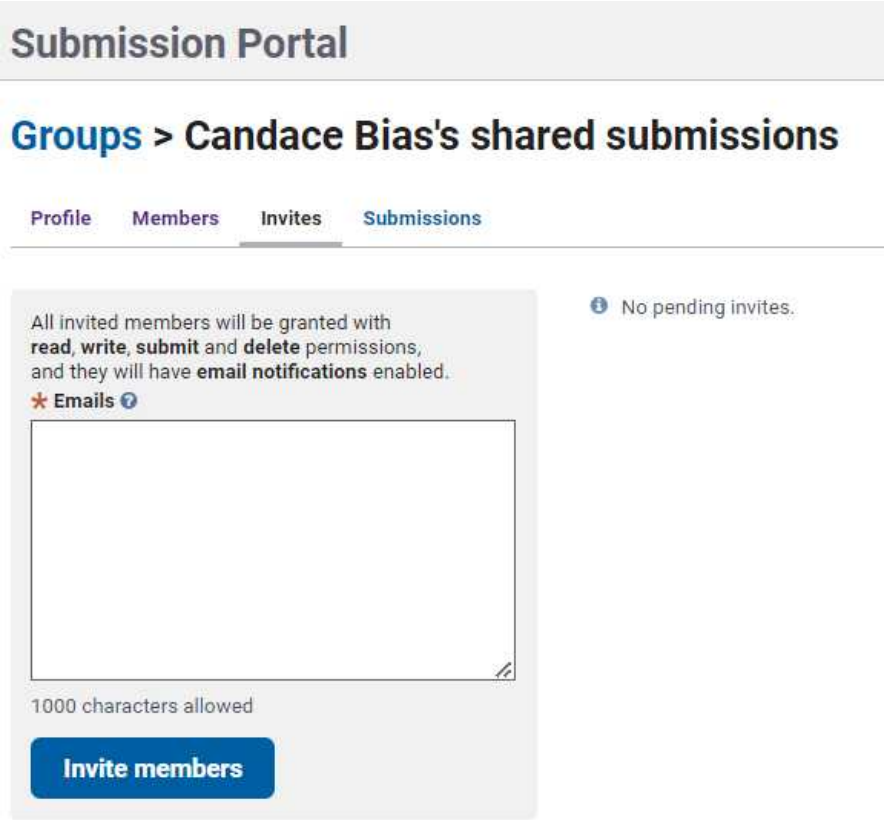
> **Note**
>
> The "Create group for shared submissions" button will not exist, if the user has not filled in all of the required profile information, marked with an asterisk ('*') on the profile page.



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.



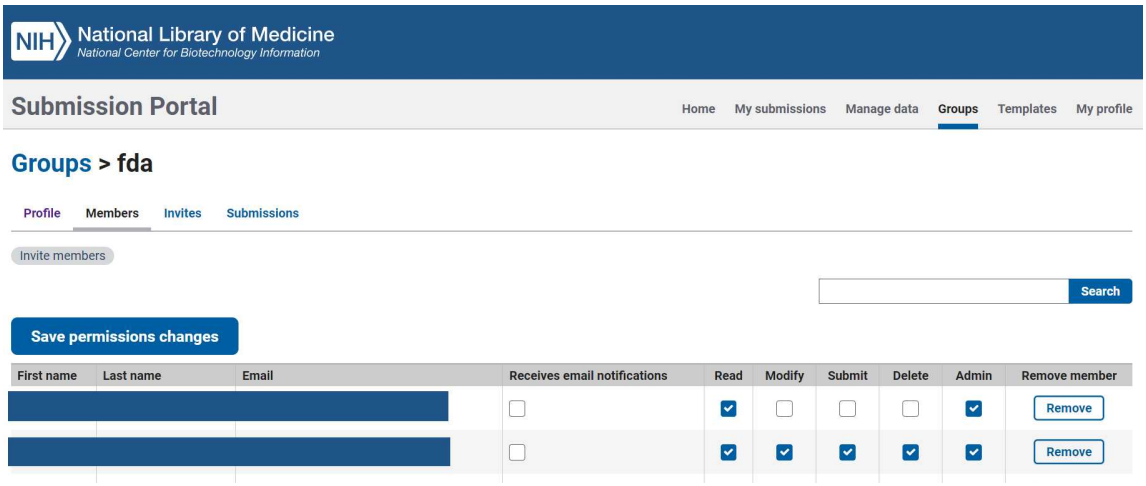
3. To invite members, use either the "**Invite members**" button at the top of the next, "Members" tab or from the "Invites" tab itself to navigate to the invite tab and add the invitees' emails to the text box. Click the "**Invite Members**" button when finished.



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.



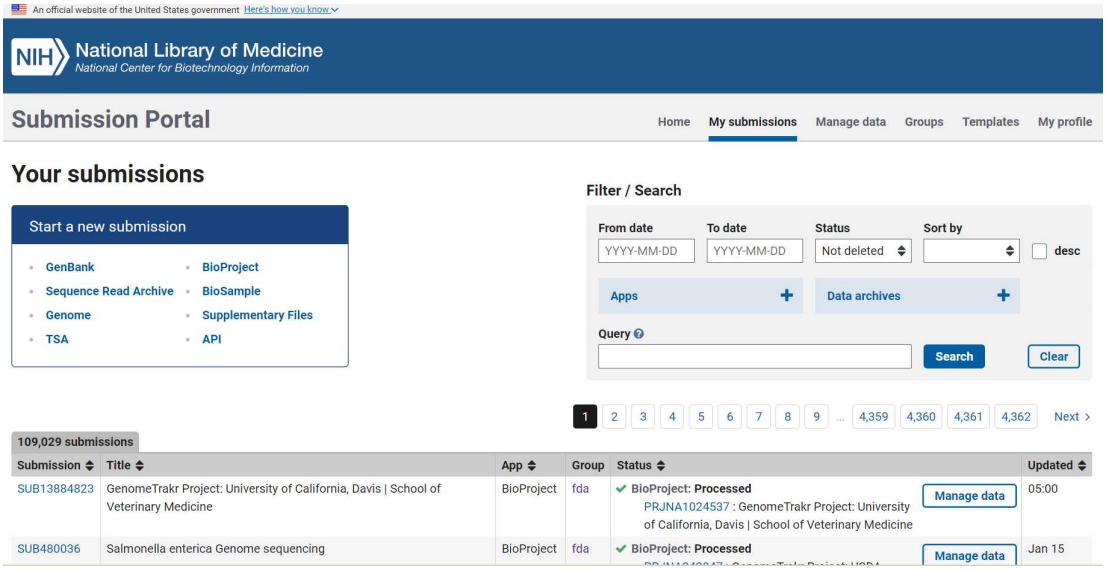
This user list should be kept current as members/staff enter and leave the laboratory.

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:

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.



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.

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.**

1.6 **Download and populate metadata templates**

Link to **VERY DRAFT** BioSample template (suggested picklist terms are in separate tab "vocabulary"):

OneHealth-Milk_v0.1.xlsx

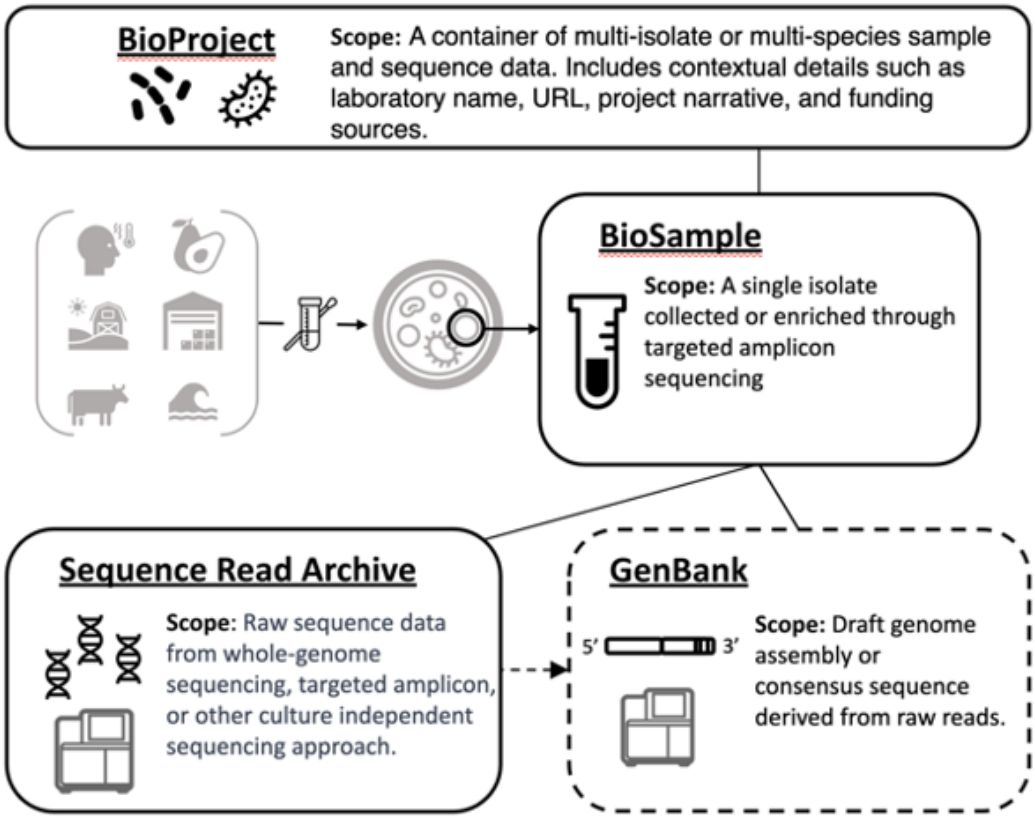
Link to **VERY DRAFT** SRA metadata template (picklist terms need to be added for H5N1 methods):

SRA-metadata-custom_Jun26.xlsx

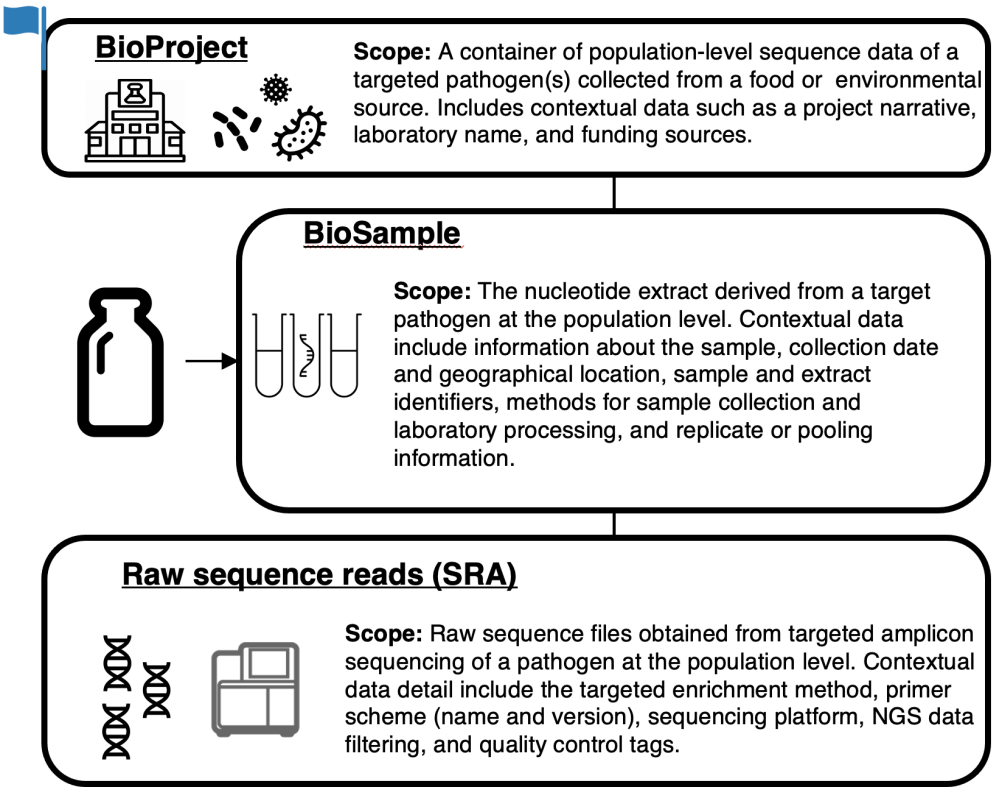
These templates are at the starting draft stage. Feel free to comment here for adding new picklist terms and/or attributes.

Data submission (BioSample and SRA)

2 **Data submission (Sample metadata, SRA metadata, and raw sequence data), compliant with the Pathogen DOM data structure.**



For single isolate or expectation of pathogen clonality within the sample. For example, milk samples on-farm, collected from an animal or from a bulk tank.



Environmental Pathogen DOM for samples with reasonable expectation of population-level sequence data. Consensus sequences not expected from these data. For example, milk sampled from silos or from Retail Milk cartons.

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.

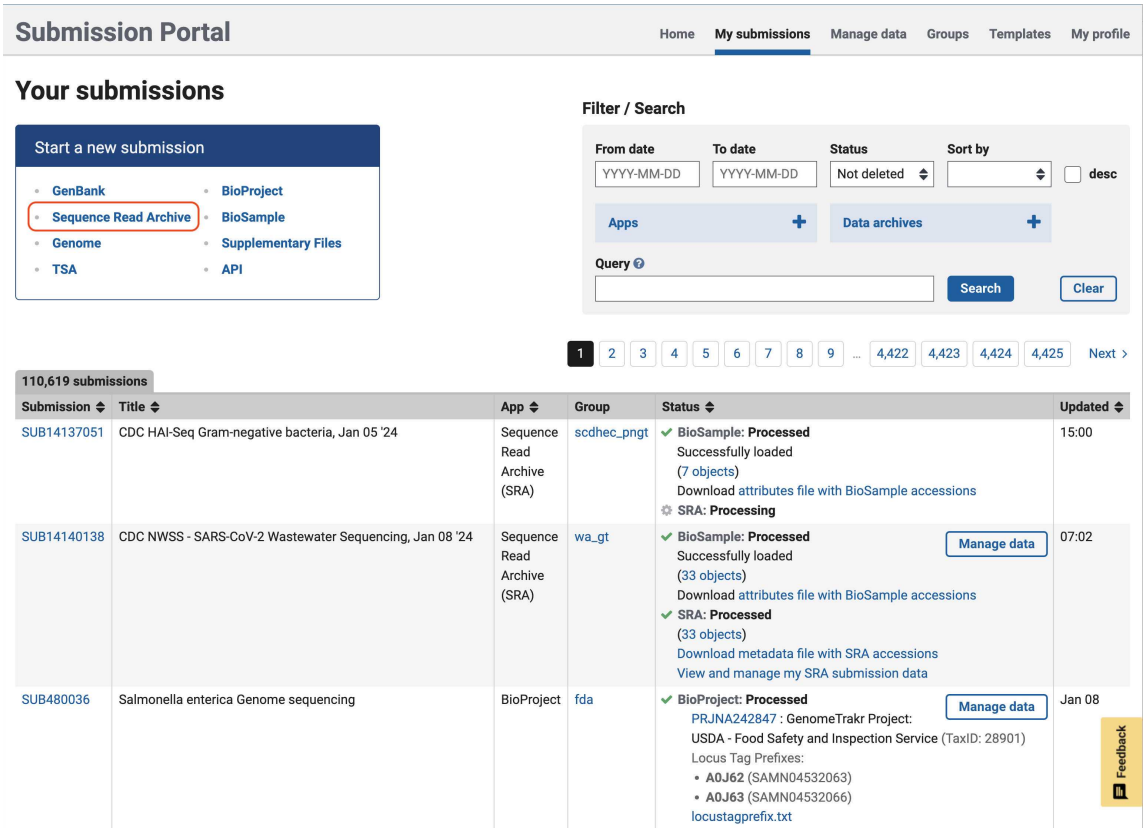
2.1 ****Pre-Submission Data Quality Control****

Verify that your sequence data meet the established quality control (QC) thresholds specific to your surveillance network.

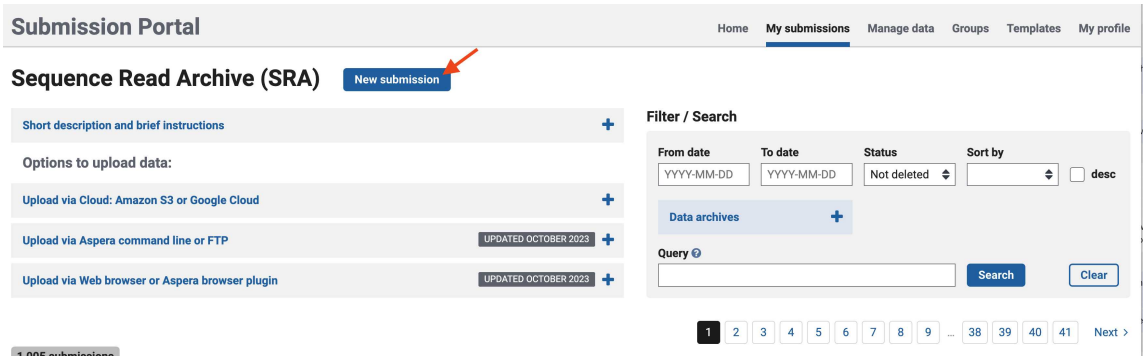
<placeholder for QC check protocol>

2.2 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.



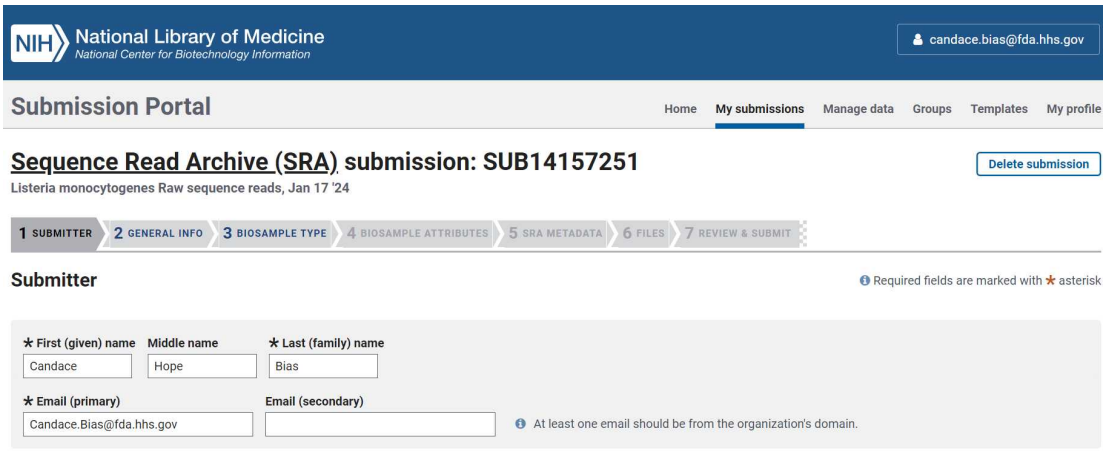
2.3 **Click** the **"New submission"** button.



2.4 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.



2.5 GENERAL INFO tab:

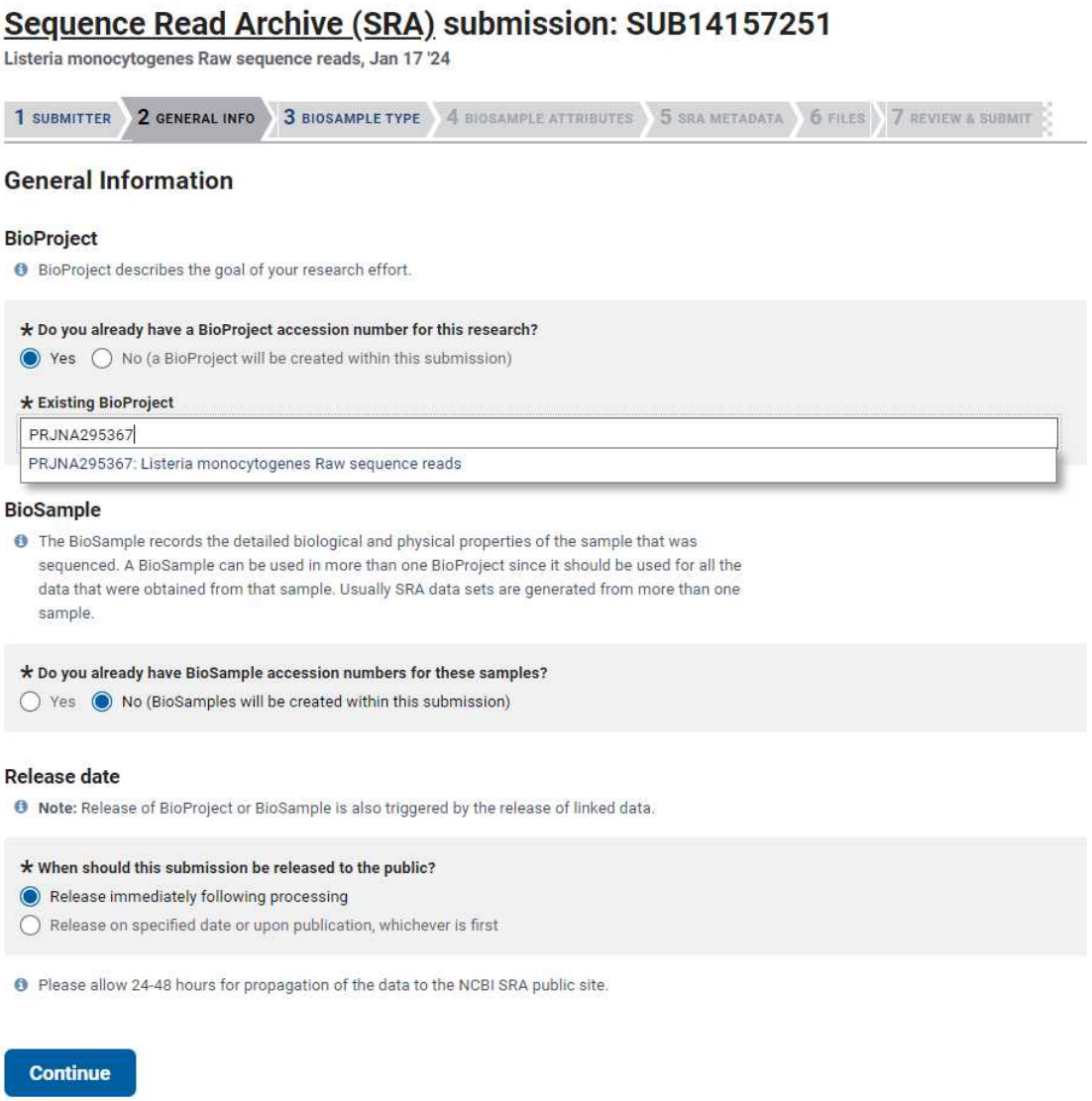
1. BioProject: Did you already have a data BioProject for this effort? If not please follow instructions in **Step 3** for creating a new data or umbrella BioProject. Return back to this sub-step with the **data BioProject** accession in hand.

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**.



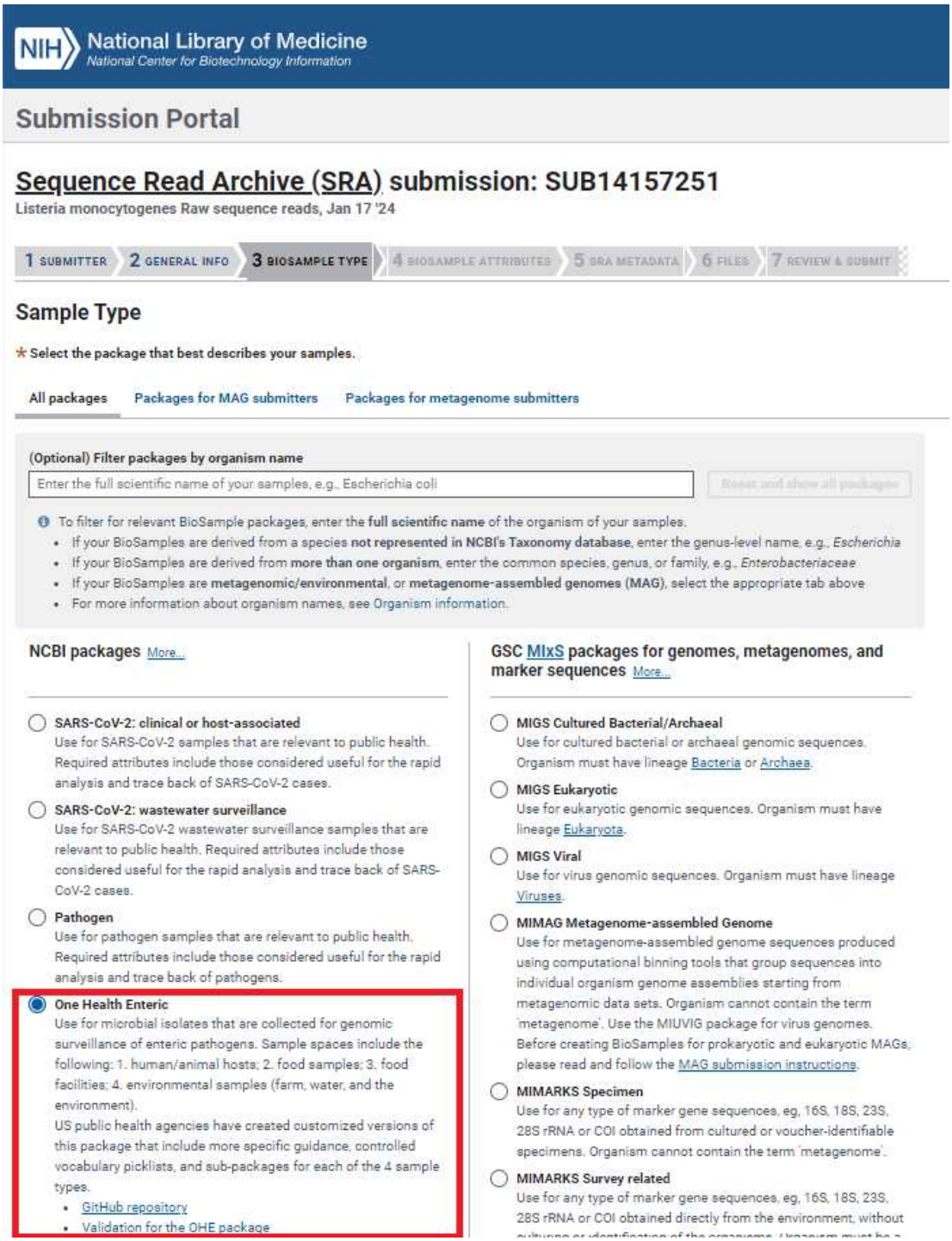
Example of filled in "General Info" tab. Please use the BioProject accession necessary for your organism and project.

2.6 BIOSAMPLE TYPE tab:

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

We recommend using the **One Heath Enteric** package for milk samples.

go to step #1.6



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

2.7 **BIOSAMPLE ATTRIBUTES** tab:

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

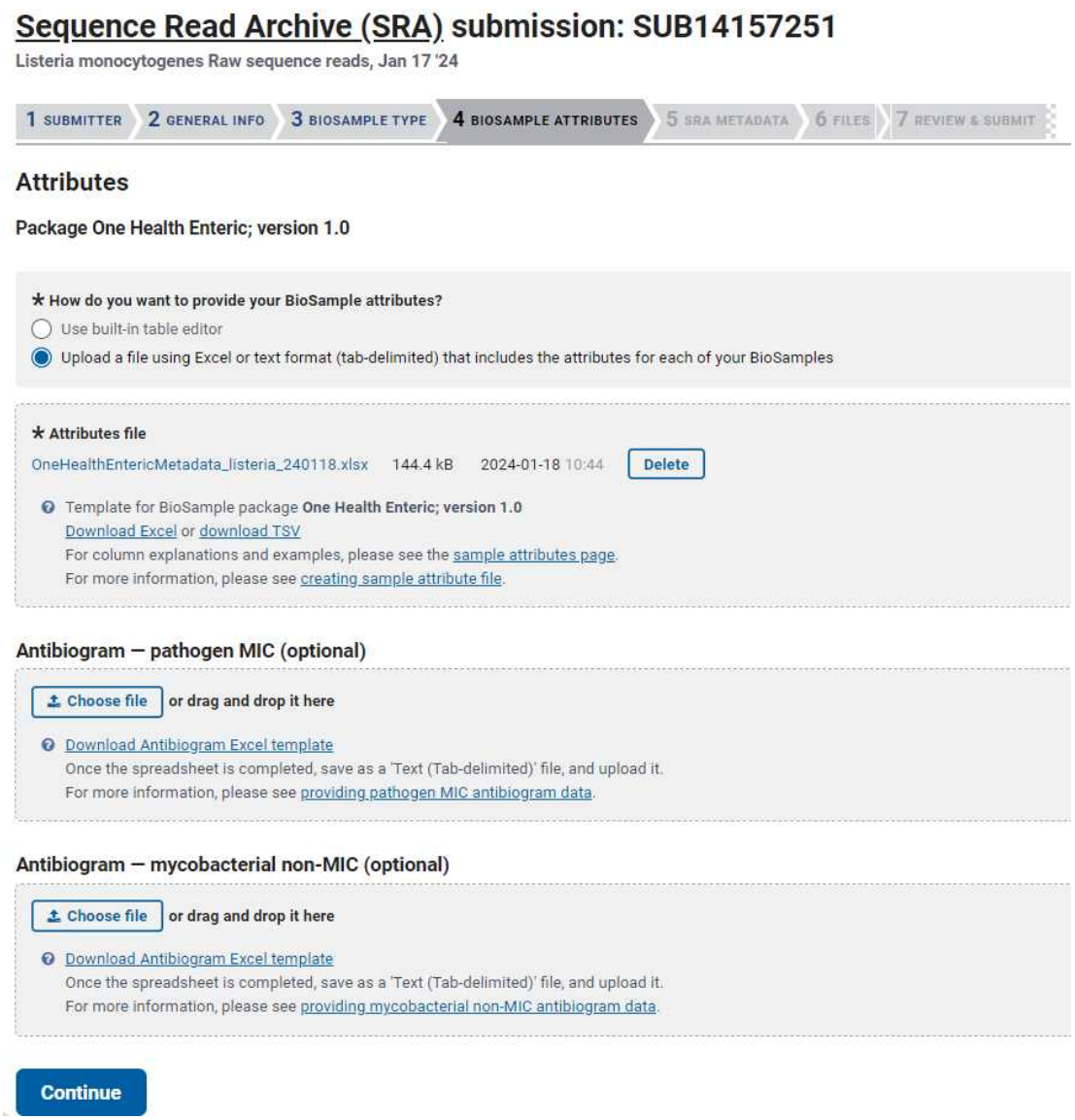
Then click "Choose File" and browse to your populated metadata template.

Note

If you have not yet populated and validated your GenomeTrakr BioSample metadata template,

Antibiogram data: please provide if you have it!

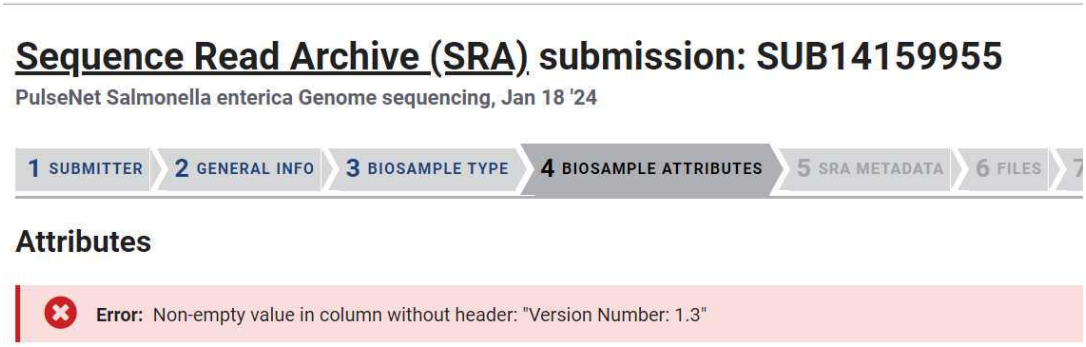
Click "**Continue**".



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.



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.

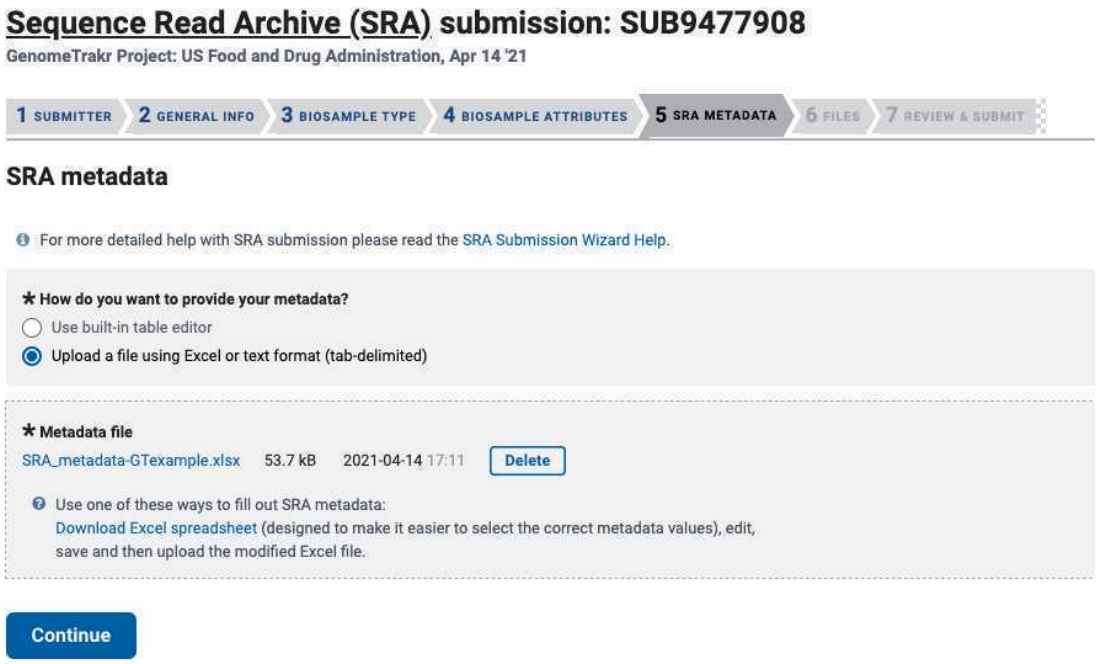
Warning: Provided taxonomy information was revised according to NCBI Taxonomy database rules. Please contact biosamplehelp@ncbi.nlm.nih.gov if you have any questions.

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**".

2.8 **SRA METADATA tab:**

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



Upload your populated **SRA metadata template** (go to step #1.6 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**".

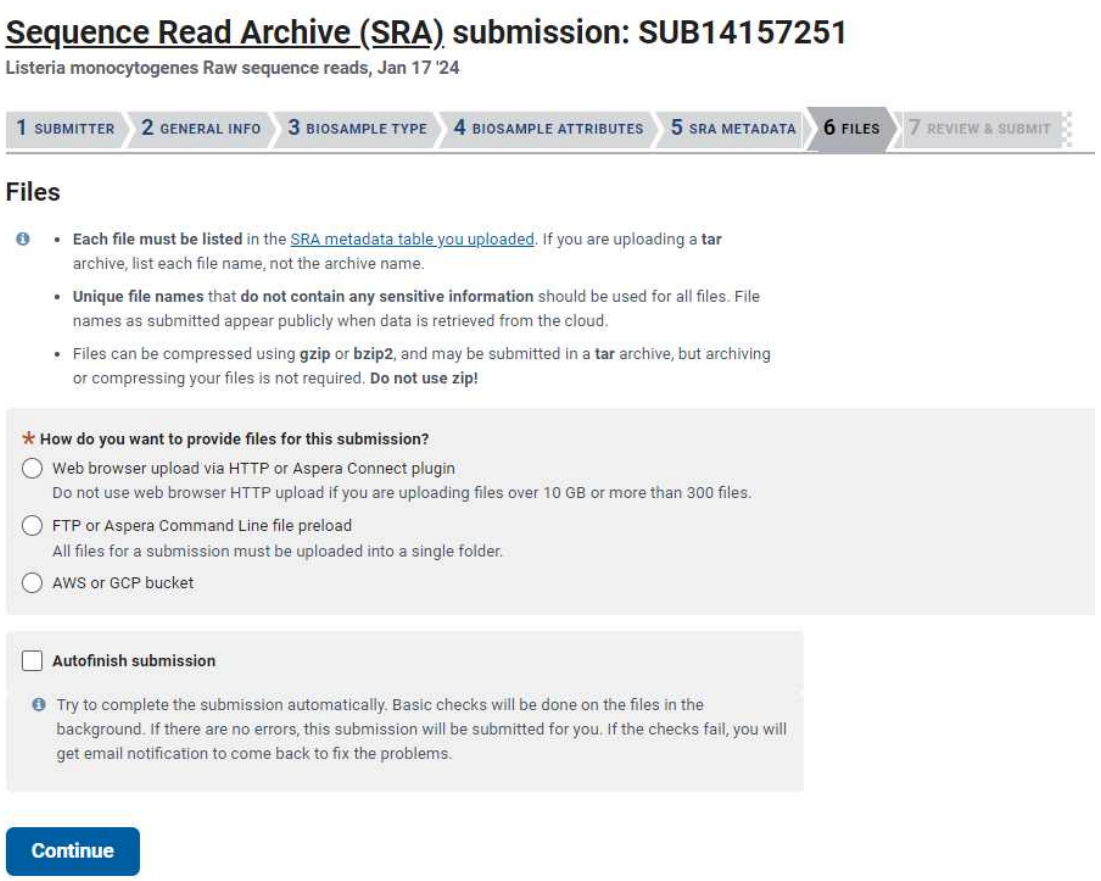
2.9 **Files tab:**

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.



2.10 **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.

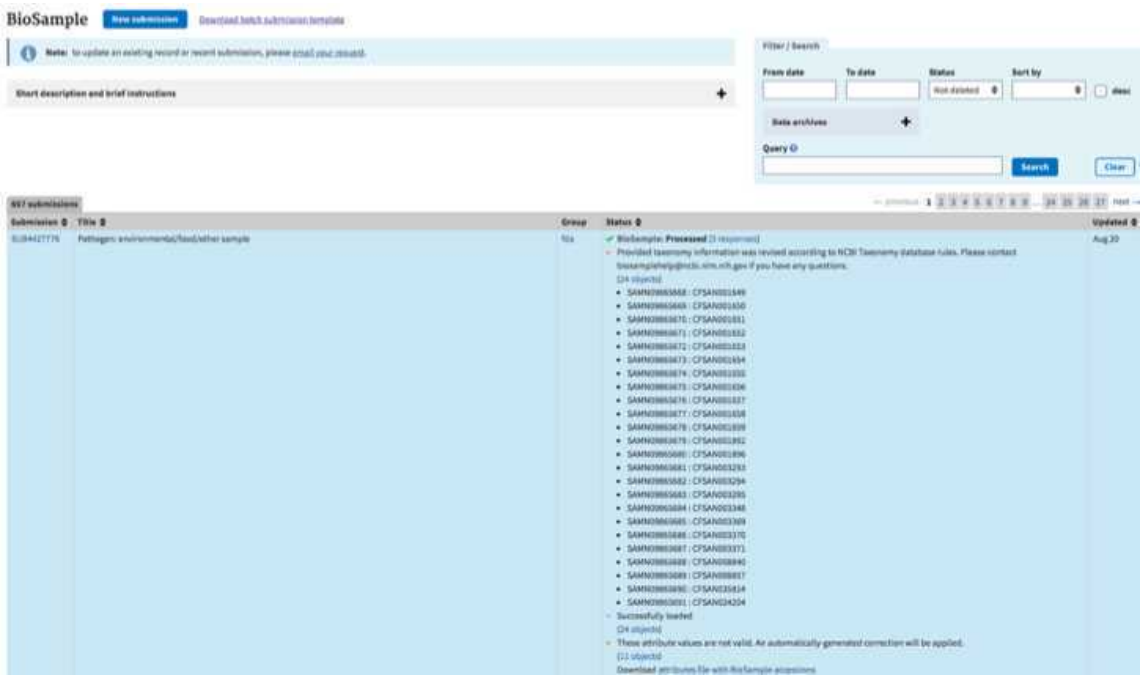
Note

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

2.11 BioSample accessions:

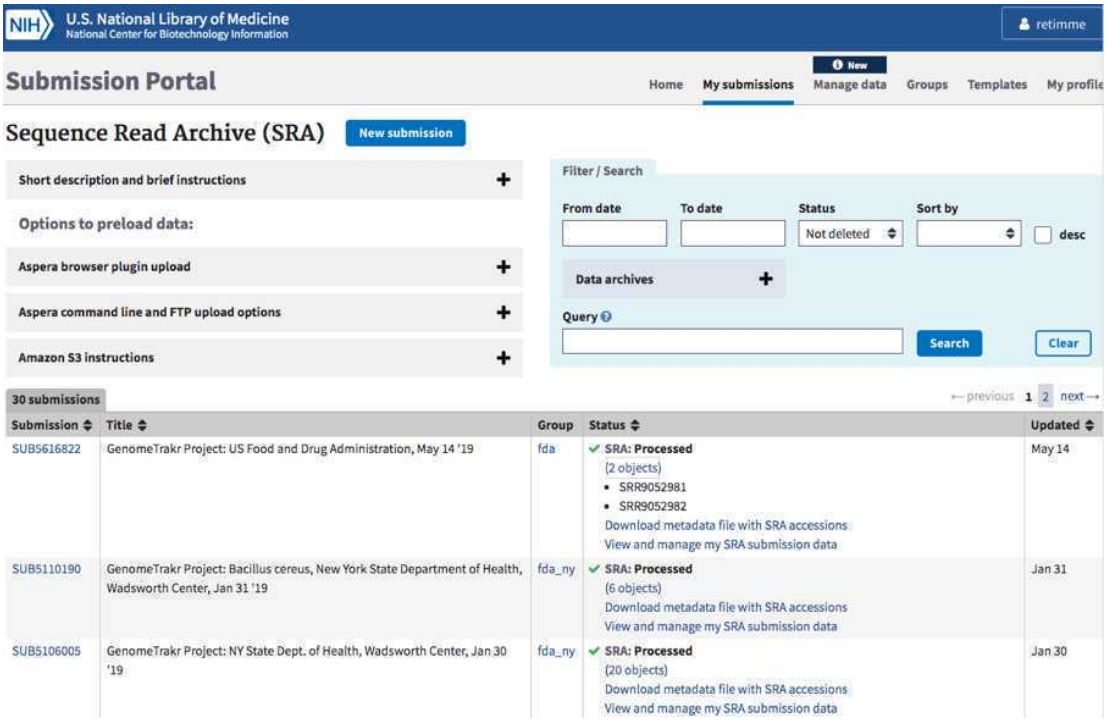
BioSample accessions will be automatically created upon submission and will be available on the “my submissions” page of the Submission portal by clicking on “## objects” within the submission record. You can also download by clicking the “Download attributes file with BioSample accessions”. Accessions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



2.12 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 SRRxxxxxxx.” You

will also receive an email with these same accessions within 24 hours, but typically much faster, containing these same accessions.



2.13 **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, some edits can be made within the submission portal and others need to be done via email.

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

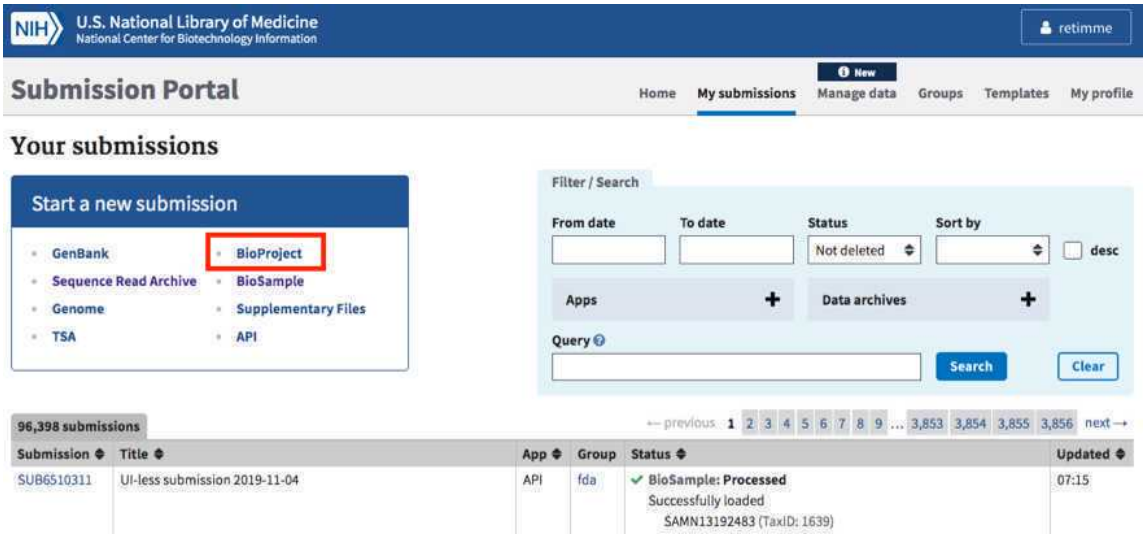
3 **Create a new BioProject**

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.

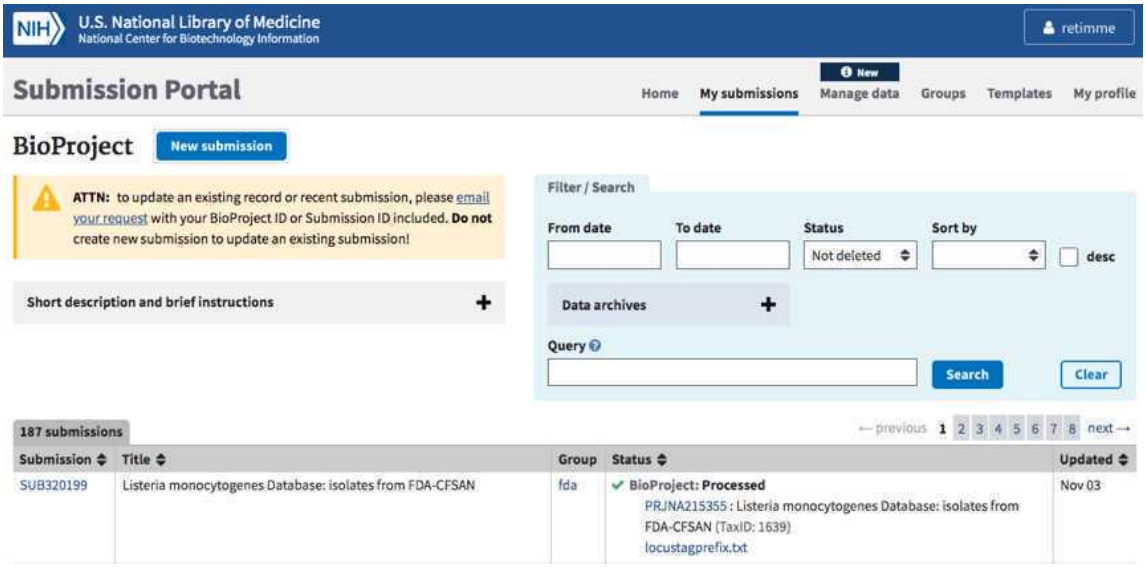
This portion of the protocol describes the steps for creating a new data BioProject linked to an existing umbrella BioProject (usually established by a coordinating group, e.g. GenomeTrakr, NARMS, Vet-LIRN).

*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.



3.2 **Click the "New submission" box:**



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 = "**Viruses**"

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., “Targeted amplicon sequencing of Influenza A (H5N1) as part of XXXX surveillance or research 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:***

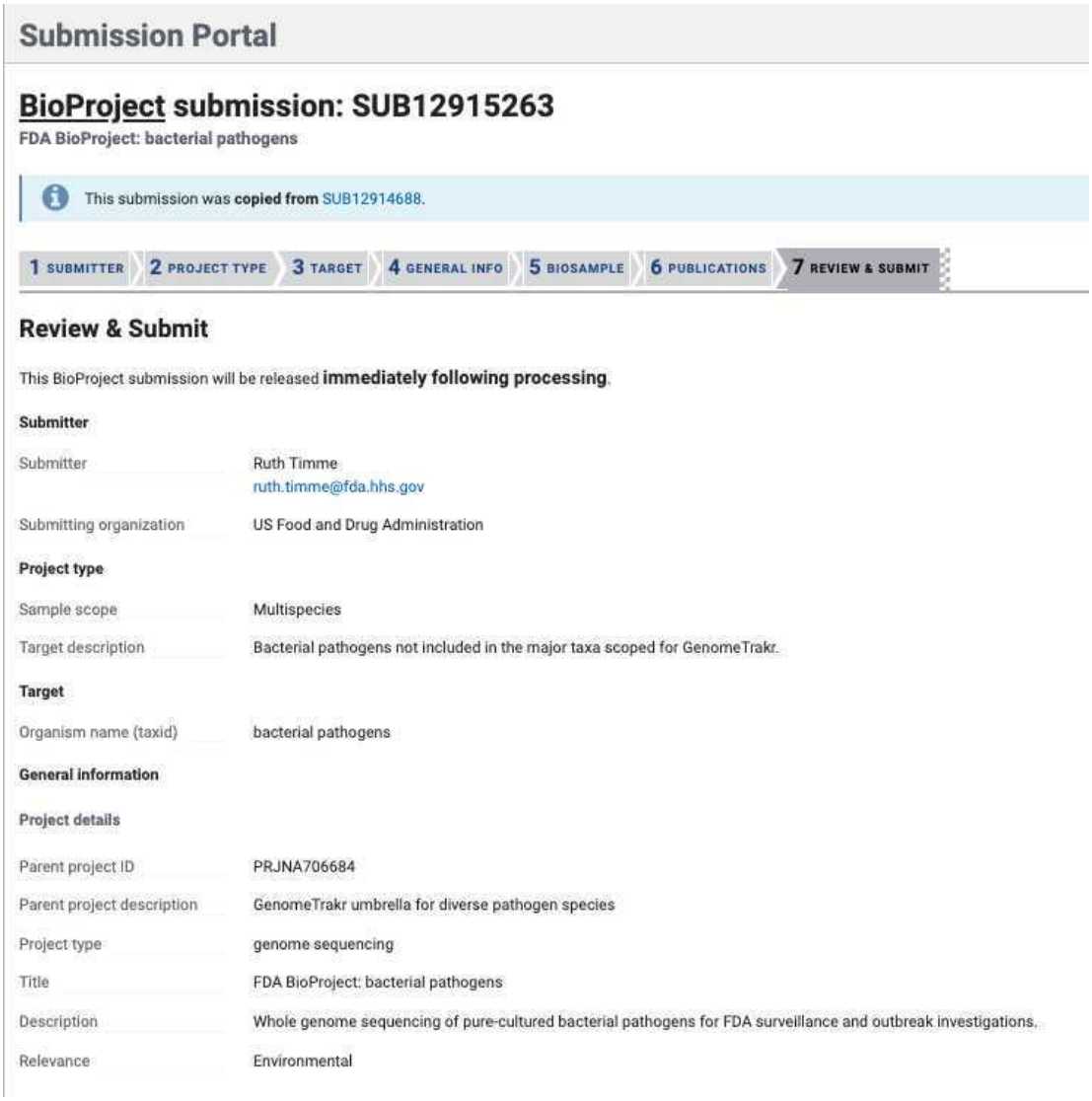
Leave blank!! You will create biosamples separately.

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 GenomeTrakr.

Target

Organism name (taxid)	bacterial pathogens

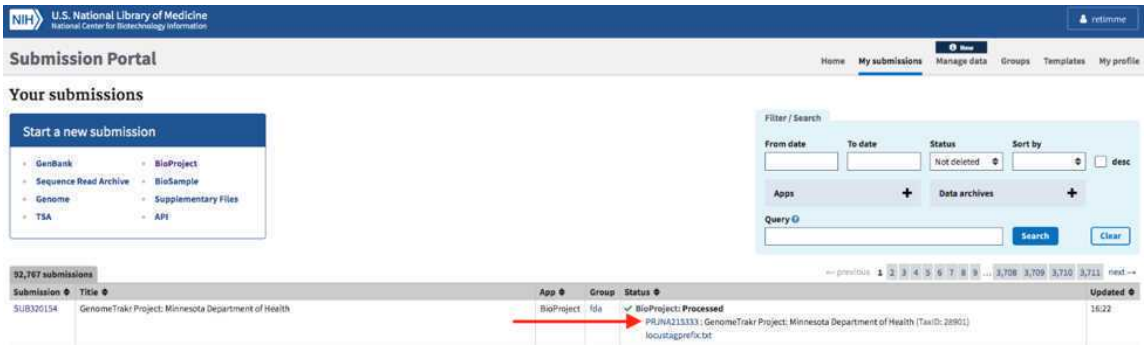
General information

Project details

Parent project ID	PRJNA706684
Parent project description	GenomeTrakr 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.



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

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**:

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.