Uploading GEOME Data to the NCBI Short Read Archive

Using GEOME for processing your metadata before submission to SRA ensures your metadata is consistent and complete, aligning with community standards as well as with other projects in your GEOME Team. GEOME's data validation steps ensure errors are caught early in the submission process.

In GEOME: Load your metadata and attach FASTQ Metadata:

You will need to upload FASTQ data into your project.

1. Goto Load Data -> Select Workbook and load your metadata
2. Goto Load Data -> Select FASTQ and following the instructions

In GEOME: Fetching your loaded Data

After you load your data into GEOME using the above steps, download your bundle to your hard drive and extract your files.

1. Goto your project, and select the download arrow on the right side of the screen next to an expedition which contains your FASTQ metadata.

NMFS_FISHES_NovaSeq-03	293	382	384	0	384	https://n2t.net/ark:/21547/FDW2	🔍 ⬇

2. Download your files and unzip the files.

Loading your data to NCBI SRA

1. **NCBI Account**

   ○ If you don't have one, create an NCBI account.

   ○ If you do, log in at the top-right of NCBI.

Step by Step Guide

1. Start a New Submission

Go to:
`https://submit.ncbi.nlm.nih.gov/subs/sra/`
Click **New submission**
forum.qiime2.org+11ncbi.nlm.nih.gov+11youtube.com+11.

---

2. Submitter Info

- Confirm your contact details.

---

3. General Info

- **Have a BioProject?** → Most will answer **No**.

- **Have BioSamples?** → Most will answer **No**.
  Answering "No" lets SRA create these records for you
  ncbi.nlm.nih.govbioinformaticsworkbook.orgncbi.nlm.nih.gov+8ncbi.nlm.nih.gov+8biostars.org+8.

---

4. Project Info

- **Project Title**: a clear short name (e.g., `DIPNet Reef Fish Sequencing`)

- **Description**: e.g., "RADSeq data for reef fish collections."

- **Relevance/Grants**: select as applicable.

---

5. BioSample Type

- Choose appropriate package:

  - "Invertebrates" or

  - "Model organism or animal sample"

    [biostars.org+11ncbi.nlm.nih.gov+11docs.hpc.oregonstate.edu+11biostars.org+1bioinformaticsworkbook.org+1ncbi.nlm.nih.gov+1ncbi.nlm.nih.gov+1](biostars.org)

---

6. BioSample Attributes

- Select **Upload a tab-delimited file**.

- Upload your `bioSample-attributes.tsv` (from GEOME)

---

7. SRA Metadata

- Choose **Upload a file**.

- Upload `sra-metadata.tsv` (from GEOME).

---

8. Upload Data Files

- You'll be prompted to install **Aspera Connect** for fast transfers.

- Choose **Fastq** or **Directory** upload, locate your FASTQ data folder.

- Use Aspera or browser to upload ZIP or `.tar.gz` files

---

9. Review & Submit

- Review all information in the **Overview** tab.

- Submit.
  → You'll receive a Submission ID (e.g., SUB123456).

---

10. After Submission

- Monitor accessions under **My Submissions** or via
  https://submit.ncbi.nlm.nih.gov/subs/sra/SUB#/overview
  ncbi.nlm.nih.gov+5ncbi.nlm.nih.gov+5youtube.com+5.

- SRA staff will process your submission and send accession numbers (e.g., SRR, SRX, SAMN).

- You'll receive emails on completion or issues.

🛠️ Troubleshooting Tips

- Use **only one BioSample** per physical sample; multiple FASTQ may map to different Experiments within the same BioSample
  en.wikipedia.org+8ncbi.nlm.nih.gov+8biostars.org+8.

- Make sure your `.tsv` headers match required SRA templates exactly.

- Average FASTQ file size <100 GB; if larger, compress or split them.

- For >5 TB submissions, break into multiple batches under the same BioProject
  protocols.ioncbi.nlm.nih.gov+2ncbi.nlm.nih.gov+2biostars.org+2.

- Unique sample names are required—no duplicates biostars.org.