

Website User Guide

Version 1.1

June 5, 2024

Contents

Section Page

[1. Introduction 1](#_Toc168499011)

[2. Manage Account 1](#_Toc168499012)

[2.1 Requesting Access to mapMECFS 1](#_Toc168499013)

[2.2 Joining an Existing Organization 1](#_Toc168499014)

[2.3 Resetting Your Password 2](#_Toc168499015)

[3. Website Structure and Terminology 2](#_Toc168499016)

[3.1 User 2](#_Toc168499017)

[3.2 Organization 3](#_Toc168499018)

[3.3 Dataset 3](#_Toc168499019)

[3.4 Group 3](#_Toc168499020)

[4. Explore Data 3](#_Toc168499021)

[4.1 Search Bar 3](#_Toc168499022)

[4.2 Results File Explorer 4](#_Toc168499023)

[4.3 Integration Tool 4](#_Toc168499024)

[5. Upload Data 4](#_Toc168499025)

[5.1 Get Permission to Upload Data 4](#_Toc168499026)

[5.2 How to Upload Data 5](#_Toc168499027)

[5.3 How to Share Data 6](#_Toc168499028)

[6. Process Data 7](#_Toc168499029)

[6.1 Tagging 7](#_Toc168499030)

[6.1.1 Recognized Data Types for Tagging 7](#_Toc168499031)

[6.2 Summary Statistics 8](#_Toc168499032)

[7. Citing mapMECFS 8](#_Toc168499033)

[8. Additional Resources 9](#_Toc168499034)

[8.1 API 9](#_Toc168499035)

[8.2 searchMECFS 9](#_Toc168499036)

# Introduction

mapMECFS is an interactive data portal providing access to research results across many biological disciplines from studies that are focused on advancing our understanding of Myalgic Encephalomyelitis / Chronic Fatigue Syndrome (ME/CFS). The goal of mapMECFS is to enable exploration of complex experimental results and connect datasets to offer new opportunities for discovery for ME/CFS researchers.

mapMECFS was created by the ME/CFS Research Network with funding from multiple NIH Institutes and the Trans-NIH ME/CFS Working Group and is hosted by RTI International.

# Manage Account

## Requesting Access to mapMECFS

mapMECFS is open to any researcher who will use the data for research purposes and is able to comply with [the Data Use Agreement (DUA](https://mapmecfs.org/files/mapMECFS_DUA_26jan2021.pdf)).

To register for mapMECFS, click **Register** in the top right corner of the home screen.



Provide the requested information, including a brief description of how you would like to use the system, and agree to the mapMECFS DUA terms. Your registration will be sent to the NIH for review and you will be notified when your account has been approved. The approval process should be quick (typically less than 2 days) and you will be notified of any delays.



## Joining an Existing Organization

Joining an existing Organization allows users to view private datasets within that Organization, which are unavailable to users outside of that Organization. Joining an existing Organization may be desirable for users who are members of a research team that use mapMECFS to collaborate and share data.

Once a new user is approved, they are added to a user-specific Organization that is created by the mapMECFS support team. A written email request to the mapMECFS support team is required to join an existing Organization.

The mapMECFS support team will then obtain approval from the Organization’s maintainer. Once approval has been obtained, the user will be added to the Organization and a confirmation email will be sent to the user.

## Resetting Your Password

From the Login page, select **Forgotten your password?** within the left sidebar.



You will be prompted to enter your email address. After submitting your email address, a link with a temporary password will be emailed to your email address. When you login using this temporary password, you will be required to create a new password.

# Website Structure and Terminology

mapMECFS is structured into distinct data entities, including Users, Organizations, Datasets, and Groups, each serving specific roles in supporting user interactions and data management.

## User

A user is an individual accessing and utilizing mapMECFS to explore, analyze, and gain insights from various datasets for research purposes.

Each user account has a defined account role. These roles are:

Organization Administrator

* View public and private datasets within their Organization.
* Create datasets.
* Edit any dataset within their Organization.
* Edit member roles within their Organization.

Editor

* View public and private datasets within their Organization.
* Create datasets.
* Edit any dataset they created.

Member

* View public and private datasets within their Organization.

To view your role in existing Organizations, select "My Organizations" from your dashboard.

## Organization

An organization is a set of users who belong to the same institution, research center, or individual research lab. Users must be associated with an Organization to upload datasets.

## Dataset

A dataset is a collection of resources (such as data files, phenotype files, result files, supporting files, or website links) with a description and study-level metadata. A dataset will generally contain one data file, one phenotype file, and an unlimited number of supporting files. During upload, users must select either public or private visibility. Public datasets are available to all mapMECFS site users. Private datasets are only available to users in your Organization.

By default, all uploaded datasets will be “private” (only viewable to the uploader and other users registered with the same organization as the uploader). For more information on how to release a dataset public, reference the How to Share Data section of this manual.

## Group

A group is a collection of datasets that are all related to a common study, cohort, experiment, or publication. For example, a group designation that connects a collection of datasets containing experimental results from multiple assays that were all run on samples derived from the same set of study participants. Datasets in a group may originate from different Organizations or the same Organization.

Note: Public and private visibility of the data still applies when a dataset is part of a group.

# Explore Data

mapMECFS offers several useful tools for finding, exploring, and viewing public datasets.

## Search Bar

The mapMECFS Search Bar function is designed to recognize user-specified terms describing multiple aspects of a Dataset. This flexibility allows users to enter keywords that identify:

* Experiments involving of a specific sample type (e.g., “blood” or “PBMCs”)
* Results of a particular experimental class (e.g., “microbiome,” “metabolomics,” or “RNAseq”)
* Studies with participants that share a specific ME/CFS case definition (e.g., “Fukuda”)
* Data files containing specific common analytes (e.g., “glucose”, “IL17”, or “EBI2”)



Note regarding synonymous terms: Because different labs apply different conventions for annotating data, it is necessary to account for synonymous terminology describing common data features. For each dataset uploaded to mapMECFS, all features (including synonymous terms) of recognized data types are automatically tagged upon upload.

More information on tagging can be found in the Tagging section of this manual.

## Results File Explorer

The Results File Explorer tool enables the search and view of specific molecules compiled across all uploaded datasets. This tool is available under the **Explorer** tab.



Using this tool, users can quickly evaluate the robustness of a given result across multiple studies and identify datasets for subsequent integrated analyses.

The results contain three separate tables:

* Data Files and Calculated Summary Statistics contains search results only from the mapMECFS-calculated summary statistics.
* Results Files contains search results only from the user-uploaded results files.
* Other Datasets contains search results from other elements of the dataset, including the title, description, and metadata.

## Integration Tool

The Integration Tool allows users to effortlessly merge clinical data files within specific cohorts. Users can specify variables of interest across several datasets to create customized, exportable datasets tailored to their specific research questions. Detailed instructions on how to use the Integration Tool can be found within the **Integration** tab.



# Upload Data

## Get Permission to Upload Data

The mapMECFS Data Access & Quality Committee (DAQ) approves submission of data to mapMECFS. The DAQ will review the Data Submission Request and decide whether to permit the submission based on the expectations outlined in the mapMECFS policy.

If submissions raise concerns related to privacy and confidentiality, risks to populations or groups, or other concerns, the DAQ will consult with other experts as appropriate. In unusual circumstances, when people are denied approval to submit data, appeals may be sent to the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke (NINDS) staff representative to the DAQ.

Steps to Request to Submit Data to mapMECFS by Data Submitters:

1. Review the capabilities of mapMECFS at https://www.mapmecfs.org/about
2. Contact the DAQ through mapMECFS (mapmecfs@rti.org) to set up an introductory meeting (preferably 2 months before the desired submission date) to discuss the dataset submission request and resolve technical and other issues.
3. The NINDS and mapMECFS staff will discuss with investigators:
	1. dataset submission expectations,
	2. dataset sharing expectations,
	3. supporting materials submission expectations,
	4. dataset access preferences,
	5. technical specifications,
	6. usage of common data elements and globally unique identifiers (GUIDs),
	7. deidentification expectations, and
	8. dataset accuracy as it relates to ME/CFS.
4. Submit a [Dataset Submission Agreement (SA):](https://mapmecfs.org/files/mapMECFS_Data_Submission_Request_21Feb2024.docx) Read the mapMECFS SA form. Submit the SA and Submitter Information and Certifications signed by all necessary parties by emailing a signed copy of this form to mapMECFS (mapmecfs@rti.org). The DAQ may determine that a submitter is exempt from an SA if the data are fully included within a peer-reviewed manuscript.
5. Dataset SA review: NINDS and mapMECFS staff will review requests to submit datasets to the mapMECFS data platform. Reviews are generally completed within 10 business days.

## How to Upload Data

Once users have obtained approval to submit data, they can upload data files in a variety of file types (TSV, TXT, PDF, etc.) or hyperlinks.

To begin uploading data, select the **Upload Data** iconin the center of the home screen.



When uploading data, you will be taken through the following upload steps:



Attempt to include as much information and metadata as possible when creating the dataset, this will help other users on the site understand the data you have uploaded.

Files should be uploaded in the expected format and a dataset should contain the following:

* One **Data File** containing sample-level values as columns and molecules as rows. For example:
	+ Gene expression counts
		- Methylation signal intensities
		- Metabolomics MS peak height

Note: An exception is that the data type Demographic, Health, and Survey (DHS) does not support a Data File format. Instead, DHS data are uploaded as a Phenotype File, as shown in the expected format example. The DHS data type is best defined as data collected from study participants that are not derived from biological samples such as blood or saliva.

* One **Phenotype File** containing subject-level clinical values, such as case-control status (“Phenotype”), age, sex, any relevant covariates, etc.
* As many **Results Files** as needed, including summary statistics for your own analysis on molecules with at least a p-value and adjusted p-value reported, as shown in this example.
* As many **Supporting Files** as needed, such as
	+ An SOP form (see this template) describing the dataset generation in more detail. Completing the SOP form is strongly recommended, as users can better understand the experimental conditions under which the Results Files were generated, as shown in this example.
	+ A link to existing publication(s)
	+ A Data Dictionary File containing information about elements in the Phenotype and/or Data Files. Please refer to the provided example (page 9), which outlines the suggested column names. Acceptable file extensions for a formatted Data Dictionary include .txt, .tsv, .csv, and .xlsx. A Data Dictionary that does not meet the mapMECFS formatting requirements should still be uploaded as a Supporting Document, but it may not interact with mapMECFS search features.

During upload, Data and Phenotype Files in the expected format will be processed for tagging and summary statistics will be calculated. Results Files will only be processed for tagging when submitted in the expected format. Tagging and calculated summary statistics are not currently available for DHS datasets. More information on summary statistics and tagging can be found within the Process Data section of this manual.

## How to Share Data

We encourage researchers to share datasets with other approved mapMECFS site users when the corresponding data is part of a manuscript accepted for publication by making their datasets “public” (viewable to all approved site users). By default, all uploaded datasets will be “private” (only viewable to the uploader and other users registered with the same organization as the uploader).

To change a dataset from private to public, open the dataset you wish to share and select Manage > Set Visibility to Public.





This will notify the mapMECFS site administration of the request, who will review the dataset for breaches of the Data Use Agreement (DUA), non-scientific content, and any personally identifiable information (PII).

Note: It is the uploader’s full responsibility to ensure none of these are present, that participant privacy is fully protected, and that sharing is compliant with all other governing policies (e.g., IRB-approved protocols, embargos, etc.). If the specific DUA changes (e.g., retroactive changes in the study’s approved IRB protocols), it is the uploader’s responsibility to remove the data or request help from mapMECFS site administrators by emailing mapmecfs@rti.org.

Data uploaded to mapMECFS should never contain PII. Users are recommended to review the U.S. Department of Health and Human Services Safe Harbor Method for more information. In addition, we recommend only including metadata relevant to the study and binning variables (e.g., age 20-30 years), in datasets with few participants (e.g., <30), so that there are always more than three participants in a group no matter how the data is stratified.

mapMECFS does not support the upload of SNP-level genotypes or raw sequencing data; instead, we recommend submitting the data to dbGaP or an alternative database. To make these data findable, users can create a dataset within mapMECFS describing the study with a link to the website containing the raw data and/or study accession IDs.

# Process Data

During upload, Data and Phenotype Files in the expected format will be processed for tagging and summary statistics will be calculated.

## Tagging

Uploaded data files are processed for tagging. This tagging process expands the search space to include the contents in the original files, molecule synonyms, and related identifiers. Tagging will only work on recognized data types; however, you can contact the mapMECFS team to request a new data type.

### Recognized Data Types for Tagging

Tagging allows users to search the contents in the original source file as well as an expanded search of relevant databases (see table below). The recognized data types for tagging include:

* Cytokine Assay
* Demographic, Health, and Survey (DHS)
* Gene Expression
* Metabolomics
* Methylation
* miRNA
* Proteomics

Note: Upload is not restricted to these data types; any data files can be uploaded to mapMECFS; tags will be cleaned annually by the DMCC for consistency across studies and to eliminate redundancy.

For a detailed table describing the different data types and which database was used for tagging, navigate to the Tagging section at the bottom of the About page: <https://staging.mapmecfs.org/about>

## Summary Statistics

For recognized Data Types mapMECFS generates a Summary Statistics file to characterize how dataset measures compare between phenotype groups as annotated in the uploaded Phenotype File. A nonparametric Wilcoxon rank-sum test is used to distinguish how dataset features differ between groups within the study (e.g., between cases and controls). Summary statistics are automatically calculated for each feature in the uploaded gene expression, cytokine assay, metabolomics, miRNA, or methylation Data Files when a correctly formatted Phenotype File is uploaded to mapMECFS.

Please note that summary statistics are processed asynchronously to avoid impacting load times. Therefore, they may not be immediately available after upload as the calculations are made in the background.

Once calculations are complete, one can view the resulting Summary Stats file by opening the dataset of interest and scrolling to Summary Statistics > View Summary Statistics.

Summary columns in this file include:

* Sample sizes in each group (labeled as “count”)
* Median value for each group
* Standard deviation
* Wilcoxon rank-sum test statistic (labeled as "Ranksum stat")
* Wilcoxon rank-sum p-value (labeled as "Ranksum p-value")
* Wilcoxon rank-sum Bonferroni Corrected p-value (labeled as "Ranksum Bonf")

# Citing mapMECFS

As outlined within the DUA, users must cite the original authors indicated within each dataset when using data obtained from mapMECFS within a publication.

The use of mapMECFS should also be referenced using the following citation:

Mathur, R.\* & Carnes, M.U.\*, et al. mapMECFS: a portal to enhance data discovery across biological disciplines and collaborative sites. J Transl Med 19, 461 (2021). <https://doi.org/10.1186/s12967-021-03127-3>

\*contributed equally and are designated co-first authors

# Additional Resources

## API

mapMECFS has an API to programmatically access, upload, and browse data and metadata.

For more information, navigate to the mapMECFS-API Repository on GitHub: <https://github.com/RTIInternational/mapMECFS-API>

Within this Repository you will find examples in several programming languages (R, python, and curl) for the following activities on mapMECFS:

* List of datasets on mapMECFS
* List of datasets and their metadata on mapMECFS
* Retrieve a list of all tags used on mapMECFS
* Retrieve metadata for a particular dataset on mapMECFS
* Search a particular tag used on mapMECFS
* Search data on mapMECFS according to a specific query
* Create a dataset and upload a resource

## searchMECFS

The MECFSnet DMCC also supports searchMECFS, an interactive search tool that allows researchers to identify and request available biospecimens to conduct novel experiments. searchMECFS houses information about the available biospecimens and associated demographic and clinical data.

More information about searchMECFS can be found at <https://searchmecfs.org/>