

Ember for Clinical Notes Search

Table of contents

- [Introduction](#)
- [High level process](#)
- [Pre-requisites for installation: Information required](#)
- [Installation: Part 1 of 2](#)
- [Installation: Part 2 of 2](#)
- [Pre-requisites for Data Ingestion](#)
- [Data Ingestion](#)
- [Playing with Notes Search](#)

Introduction

With up to 80% of healthcare data being unstructured, MetiStream helps healthcare organizations gain deeper insights. Ember Clinical Notes Search (ECN) allows Quality analysts and hospital staff to exploit the data once locked away in the physician's clinical notes. Built on Ember's scalable big data and analytics platform, ECN can tackle a multi-year backlog of notes in a matter of hours. Using Natural Language Processing (NLP) and big data search tools, ECN automates the processing of these free-text fields and stores them as structured, reportable data that can be rapidly searched and analyzed.

The screenshot displays the Ember Clinical Notes Search interface. On the left, a sidebar contains search filters: 'Exact Text Search', 'All Ontology Search', and 'Advanced Search'. Under 'Queries', there are buttons for 'Myocardial Infarction', 'Colonoscopy', 'Notes mentioning colon', and 'Malignant Hyperthermia'. A 'Save Query' button is at the bottom of the sidebar. The main area shows search results for 'No Filters'. At the top, it indicates '74418 results found in 756ms'. Below this is a table with five rows of results. Each row contains a 'Mimic_100k_Benchmark ID', 'Category', 'Chart Date', 'Description', and a 'View Note' button. The table also includes a 'Unit No.', 'Admission Date', 'Discharge Date', and 'Date of Birth' for each entry. At the bottom of the table, there are navigation buttons: 'Prev', '1', '2', '3', '4', '5', and 'Next'.

Mimic_100k_Benchmark ID	Category	Chart Date	Description	View Note
24575	Discharge summary	08/08/2133 8:00 PM	Report	View Note
24675	Discharge summary	06/06/2138 8:00 PM	Report	View Note
19381	Discharge summary	03/21/2107 8:00 PM	Report	View Note
26568	Discharge summary	05/11/2185 8:00 PM	Report	View Note
28566	Discharge summary	09/01/2137 8:00 PM	Report	View Note

High level process

There are two simple steps to get Ember up and running and search your clinical data.

1. Install / configure the product in GCP
 - a. This is where you would specify few parameters / provide permissions for ember to run in GCP
2. Data Ingestion
 - a. Point ember to your CSV file and let it mine the data to enable you to do clinical search

Pre-requisites for installation: Information required

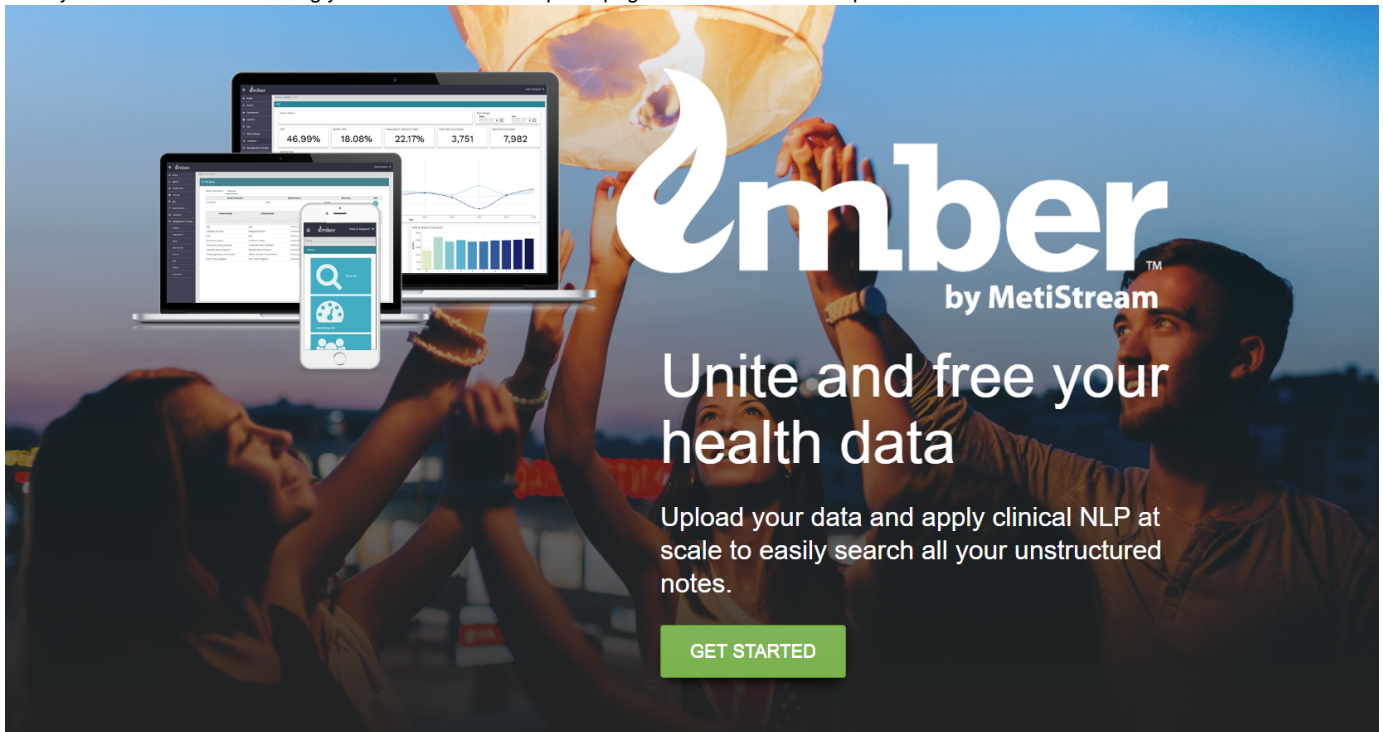
In order for Ember to be properly configured the following information is needed. If you do not know the answers to this question, please contact your IT administrator to get the answers to these questions.

- Ember License: Please use this form (<https://forms.gle/KhgWFAKu1Sj8xWzc7>) to request a license. For additional help, please contact MetiStream at onboarding@metistream.com.
- Ember needs proper access to be able to ingest data. Please set up a service account with necessary access and save the key. Ember will need you to specify *Project ID*, *Region*, *Zone*, *Service Account Name*, and *Service Account Credential(Key)* at step 6. Click here for more details: [Creating and managing service accounts \(Link\)](#)

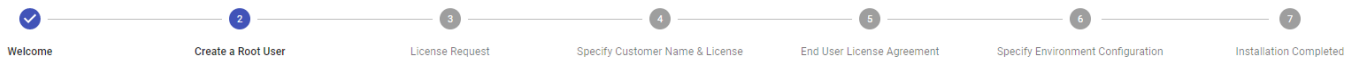
- After you create a service account, grant the necessary roles to it. Ember needs *Google Cloud Storage* and *Dataproc* access. Click here for more details: [Granting roles to service accounts \(Link\)](#)
- Create and save the key for future use. Click here for more details: [Creating and managing service account keys \(Link\)](#)
- You will need to specify region and zone so Ember knows where to launch clusters to ingest datasources. Click here for more details: [Available regions and zones \(Link\)](#)
- Ember also needs users to have a valid UMLS (Unified Medical Language System) credential to generate complete NLP contents. Please refer to the below link on how to apply for an account. Click here for more details: [How to License and Access the UMLS Data\(Link\)](#)

Installation: Part 1 of 2

Once you launch ember first thing you see would be our splash page. Click Get Started to proceed.



1. Create a user account to use to logon to setup ember



Create Account

Set up an account in Ember. You will use the *email* and *password* to log in.

The password needs to be at least 8 characters long and contains 1 uppercase, 1 lowercase letter, and 1 special character.

User ID	First Name
<input type="text"/>	<input type="text"/>
Last Name	Email
<input type="text"/>	<input type="text"/>
Password	Confirm Password
<input type="password"/>	<input type="password"/>

User account has been created successfully.

Please click **Next** if you already have a valid license.

or

Please use this form (<https://forms.gle/KhgWFAKu1Sj8xWzc7>) to request a license. For additional help, please contact MetiStream at onboarding@metistream.com

Back **Next**

Once your account is created, you can request a license by filling out the form using the link in the confirmation page to request license.

Installation: Part 2 of 2

This step can started once you have a valid license to use. Copy and paste the license you got from MetiStream and validate the license

Please fill in your **organization name or full name** and **license**. The name needs to match the one in your license request exactly.

Company Name or User Full Name
MetiStream

License Code
XXXXXXXXXXXXXXXXXXXX

Check License

Accept the user agreement.

TERMS AND CONDITIONS OF SERVICE

These terms and conditions of service are a legal agreement (the "Agreement") between you (together with the business entity, if any, that you represent, "Customer") and MetiStream, Inc., a Virginia corporation, with its principal place of business at 1953 Gallows Road, Suite 570, Vienna, VA 22182 ("MetiStream"), establishing terms and conditions under which MetiStream shall make available certain MetiStream products on an evaluation basis to Customer ("MetiStream Products"). Each of MetiStream and the Customer may be referred to herein as a "Party" or collectively the "Parties." The date Customer first agrees to or accepts this Agreement, or that Customer otherwise first accesses and uses the MetiStream products following the date this Agreement is first made, is referred to herein as the "Effective Date."

BEFORE YOU CLICK ON THE "I ACCEPT" BUTTON OR OTHERWISE INDICATE ACCEPTANCE OF THIS AGREEMENT, CAREFULLY READ THE TERMS AND CONDITIONS OF THIS AGREEMENT. BY CLICKING ON THE "I ACCEPT" BUTTON OR USING THE SERVICES, YOU AND THE BUSINESS ENTITY THAT YOU REPRESENT ARE AGREEING TO BE BOUND BY AND ARE BECOMING A PARTY TO THIS AGREEMENT AS THE "CUSTOMER." YOU REPRESENT AND WARRANT TO METISTREAM THAT YOU HAVE THE AUTHORITY TO ENTER INTO THIS AGREEMENT ON BEHALF OF THE BUSINESS ENTITY YOU REPRESENT. YOU HEREBY COVENANT TO METISTREAM TO ENSURE THAT ALL CUSTOMER'S USERS THAT GAIN ACCESS TO THE METISTREAM PRODUCTS AND/OR SERVICES ARE BOUND BY THIS AGREEMENT. IF NO ENTITY IS NAMED WHEN YOU CLICK "I ACCEPT", YOU ARE THE ONLY PERSON WHO MAY USE THE METISTREAM PRODUCTS. EACH INDIVIDUAL WHO USES THE METISTREAM PRODUCT WILL BE REQUIRED TO ENTER INTO THIS AGREEMENT OR METISTREAM'S END USER LICENSE AGREEMENT. IF YOU DO NOT AGREE TO ALL OF THE TERMS OF THIS AGREEMENT, THEN DO NOT CLICK "I ACCEPT" AND YOU WILL NOT BE PERMITTED TO ACCESS AND/OR USE THE SERVICES.

EVALUATION AGREEMENT

In consideration of mutual promises hereinafter set forth, it is agreed by and between the Parties as follows:

1 Evaluation License. Subject to the terms and conditions of this Agreement, MetiStream hereby grants to Customer a non-exclusive, revocable, and non-transferable right during the Term (a) to install and use the object code version of the MetiStream Products at third party data centers under the direction of Customer (the "Hosted Eval License"); or (b) to access the online functionality of the MetiStream Products operated by MetiStream and made available to Customer (the "Cloud Eval Services"), and together with the Hosted Eval License, the "Evaluation Products and Services" and, in each case, the related Documentation and solely for Customer's internal evaluation purposes. "Documentation" means MetiStream's on-line information and user materials relating to the use of the Evaluation Products and Services made available to Customer in connection with this Agreement.

1.1 Ownership. MetiStream and its licensors reserve all right, title and interest in and to the Evaluation Products and Services under all applicable federal, state and local laws of the United States and any other jurisdiction. Customer is not acquiring any right or interest in the Evaluation Products and Services except the access rights pursuant to Section 1.

1.2 MetiStream acknowledges that as between the parties, the Customer is and will be the exclusive owner of all right, title and interest in and to its data, including all intellectual property rights therein. Customer hereby grants to MetiStream a non-exclusive, worldwide, non-transferable, non-sublicensable, right and license to use its data during the Evaluation Period and such additional periods, if any, as MetiStream is required to perform services under this Agreement. Customer acknowledges and agrees that Customer remains solely liable and responsible for Customer's data. In no event shall MetiStream have any liability related to Customer's data.

1.3 Restrictions. Customer will not, directly or indirectly, and will ensure that no authorize user shall, (a) sublicense, transfer, or otherwise assign its rights in the Evaluation Products and Services to any third party nor allow any third party to access or use the Evaluation Products and Services; (b) modify the Evaluation Products and Services, create derivative works of the Evaluation Products and Services, or translate, reverse engineer, de-compile, or disassemble the Evaluation Products and Services to develop any other computer program or for any other reason; (c) copy the Evaluation Products and Services, in whole or in part, without the prior written consent of MetiStream or (d) make the Evaluation Products and Services available to any third party, including as a service bureau, ASP, or other service provider. The original and all copies of the Evaluation Products and Services remain the sole property of MetiStream, subject to all of the confidentiality and other restrictions set forth in this Agreement. Customer must retain all legends relating to copyright, trademarks, patents, or confidentiality on all copies of the Evaluation Products and Services.

2 Term and Termination.

2.1 Term. This Agreement, unless earlier terminated, will remain in effect for a period of ten (10) days beginning on the Effective Date. The parties may agree to extend the Evaluation Period from time to time.

Back **Accept**

Here is where you need to provide permissions to ember to run in GCP and process your clinical data.

✓

Welcome

✓

Create a Root User

✓

License Request

✓

Specify Customer Name & License

✓

End User License Agreement

6

Specify Environment Configuration

7

Installation Completed

Please provide the **UMLS credential** and **Google Cloud configuration**. If you need more information, please refer to the resources below.

Click to see more instructions on Google Cloud configuration

Click to see more instructions on UMLS credential

UMLS Credential

UMLS Username

UMLS Password

Google Cloud Configuration

Project ID

Region

Zone

Credential

Service Account

✓

Welcome

✓

Create a Root User

✓

License Request

✓

Specify Customer Name & License

✓

End User License Agreement

✓

Specify Environment Configuration

7

Installation Completed

Back

Finish

Pre-requisites for Data Ingestion

1. URL to your CSV file uploaded to the GCP storage
2. What delimeteres / characters used for the CSV file
3. Which column is the primary key and which column has thetext

Data Ingestion

As soon as the installation is done, it directly starts the data ingestion wizard to process your data.

1
2
3
4

Configure Data Set
Configure NLP Column
Configure Search and Display Fields
Review and Ingest

Set the name of the data set

Name

GCP-Test

Configure the path where the data is located (Google storage bucket)

Uploading data to Google Storage(Link)

Path

gs://tr[REDACTED]nic_100.csv

Number of notes to process (limited to 1000 for demo)

Note Count

100

Set the data format

Source Type

CSV

Set additional data format information (optional)

Back
Next

We can directly point to your CSV file in the GCP cloud storage to process your data. Give a name to your dataset and enter the URL of your CSV file from the cloud and specify the file type as CSV. If you have specific characters used in your csv file like quote or escape character like “ please specify that.

Ember scans your data and shows the schema. You can click preview if you like to view few rows of your data.

Remember to specify the primary key and the text column to do NLP.

1
2
3
4

Configure Data Set
Configure NLP Column
Configure Search and Display Fields
Review and Ingest

Set the text field to run NLP on and the primary key of the data source.

Text Field

Primary Key

Field Name	Data Type
ROW_ID	Integer
SUBJECT_ID	integer
HADM_ID	Integer
CHARTDATE	timestamp
CHARTTIME	string
STORETIME	string
CATEGORY	string
DESCRIPTION	string
CGID	string
ISERROR	string

10

1 2 >

Preview

(Optional) Click to preview sample rows from the data (Note: may timeout on large data sources)

Back
Next

Data Set Creation Wizard

1

2

3

4

Configure Data Set

Configure NLP Column

Configure Search and Display Fields

Review and Ingest

Set the text field to run NLP on and the primary key of the data source.

Text Field

Primary Key

Field Name	Data Type	Row1	Row2	Row3
ROW_ID	Integer	174	175	176
SUBJECT_ID	Integer	22532	13702	13702
HADM_ID	Integer	167853	107527	167118
CHARTDATE	timestamp	2151-08-04T00:00:00.000Z	2118-06-14T00:00:00.000Z	2119-05-25T00:00:00.000Z
CHARTTIME	string			
STORETIME	string			
CATEGORY	string	Discharge summary	Discharge summary	Discharge summary
DESCRIPTION	string	Report	Report	Report
CGID	string			
ISERROR	string			

10

12>

Preview

(Optional) Click to preview sample rows from the data (Note: may timeout on large data sources)

Back

Next

The next step is optional to specify which columns to use.

Data Set Creation Wizard

1

2

3

4

Configure Data Set

Configure NLP Column

Configure Search and Display Fields

Review and Ingest

Note: This step is optional

Specify fields to be displayed during searches ⓘ

Display Fields

Search Fields

Add Item

Add Item

Data Set Creation Wizard

1

2

3

4

Configure Data Set

Configure NLP Column

Configure Search and Display Fields

Review and Ingest

Please review the following items. Click 'Finish' to complete configuration and begin data ingestion.

Data Source Configuration

Field	Value
Name	GCP-Test
Path	gs://[REDACTED]100.csv
Data Type	CSV
Text Field	TEXT
Primary Key	ROW_ID

Job Metrics

Field	Value
Note count	100
Number of nodes	2
Estimated time (min)	~ 10 minutes
Estimated cost	\$0.37

⚠ Warning: Clicking 'Finish' will create a Dataproc cluster in your GCP environment and begin the ingestion process. This will incur costs to your project.

Back

Finish

Once you click finish data processing will start.

Playing with Notes Search

Once on the search page you can use several different fields to search for the desired record. These search fields are configurable in the Data Source settings in the Management Console

- **Exact Text Search** allows you to search all notes within the data source using normal free-text. This will look for exact text and phrase matching within the note
- **All Ontology Search** searches the notes for a specific ontology codes across multiple languages. In order to utilize this function, Ember analyzes all notes in the data source and annotates the text with the ontology codes using Ember's Natural Language Processor.

Note: The currently supported ontology Snomed, ICD-9, ICD-10, RxNorm, LOINC, HPO, HPCPS

- **Advanced Search** allows users to provide more detailed search criteria. See *Advanced Search* below for more details