

A Chatbot to Streamline Biomedical Data Discovery and Analysis

João R. Almeida

1. DETI / IEETA, University of Aveiro, Aveiro, Portugal

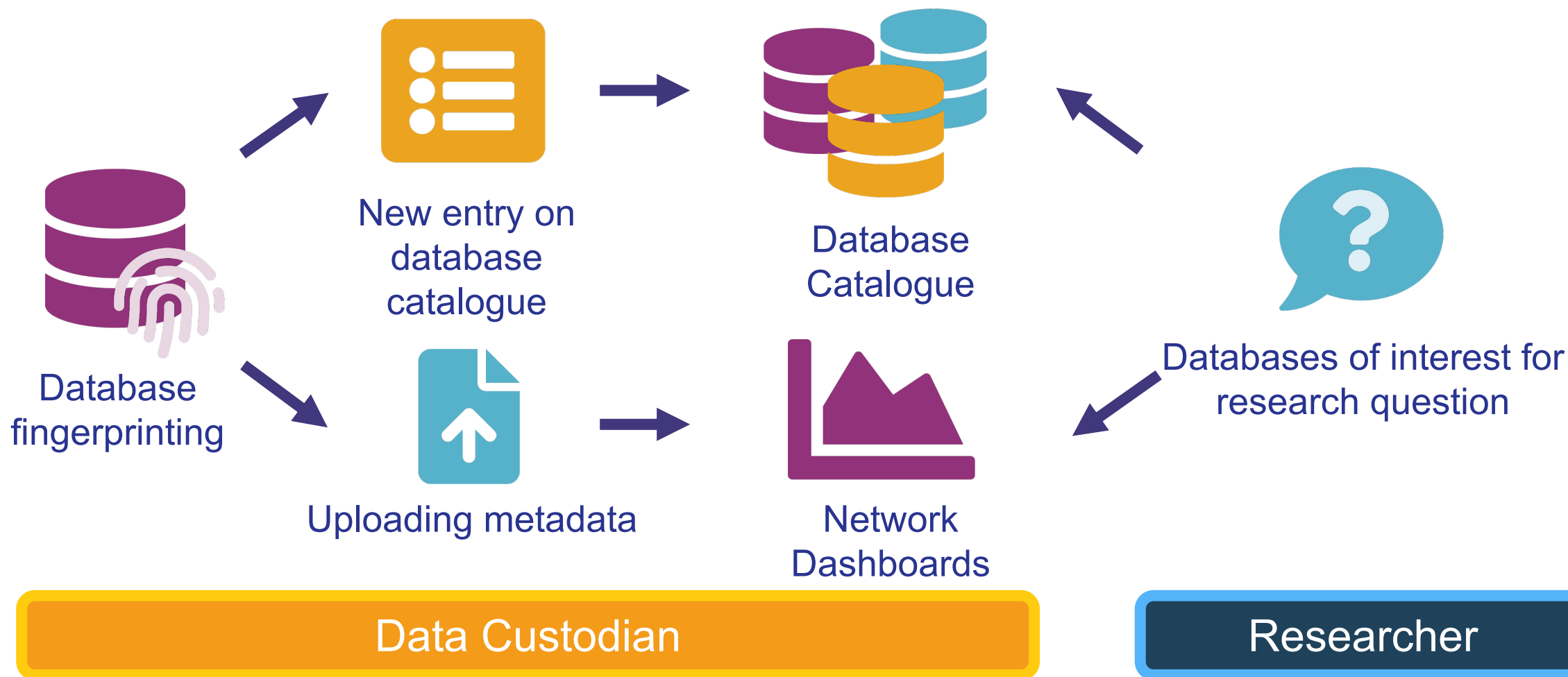


Where do we find data?





Database fingerprint



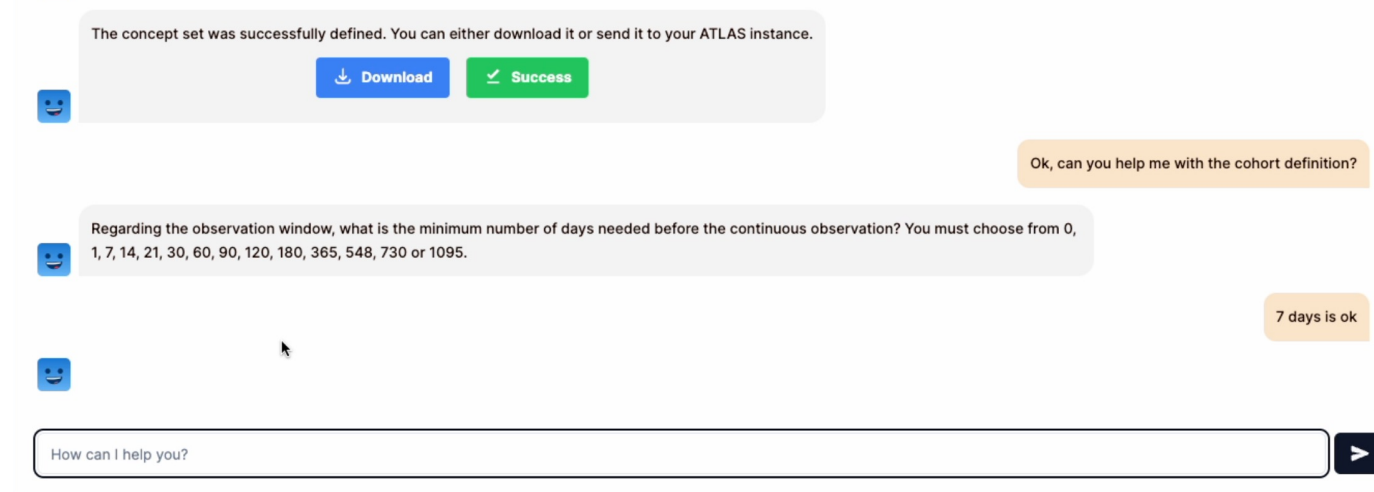


How to find the most suitable databases?



Chatbot overview

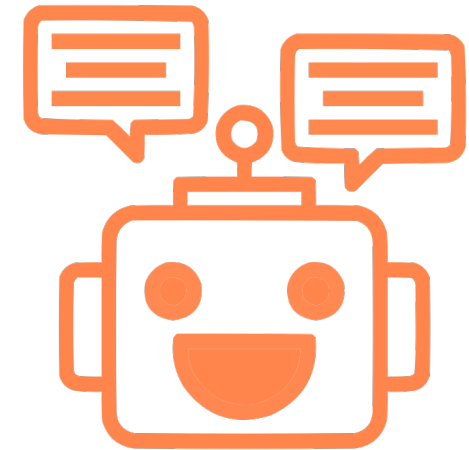
- Enable discovery and basic feasibility enquiries
 - Using EHDEN Network Dashboards data
- Compare basic characteristics
 - With other Data Partners in the network
- Provide suggestions of most suitable databases
 - Information Retrieval
 - Large Language Models
 - RAG techniques





Technology used

- Information Retrieval
 - BM-25 model
 - Used to retrieve the most suitable databases
 - Based on the information extracted from the Catalogue Export package
- Large Language Models
 - Llama3 70b
 - Used as a NLP tool
 - To extract information from user's questions
 - Also used to create a fluid conversation



Database discovery

EHDEN PORTAL Free text search EHDEN

Hippocrates

I'm here to help you find the best OMOP CDM databases for your observational studies. Please provide me with questions or specific details about your study needs so I can identify the most suitable databases for you.

Hello, could you please describe the scope of your study?

We want to conduct a characterization study with the main research topic about covid-19. Can you assist us?

Hello! Based on your interest in COVID-19 research, I recommend exploring these specialized medical databases that focus on infectious diseases and public health:

- Tampere University Hospital
- UK Biobank
- Norwegian health registries at University of Oslo

These databases will provide you with valuable resources to support your study.

Are there any other concepts you'd like to add? If yes, please add them. If no, simply respond with 'no'.

Database discovery

EHDEN PORTAL Free text search EHDEN

Hippocrates

I'm here to help you find the best OMOP CDM databases for your observational studies. Please provide me with questions or specific details about your study needs so I can identify the most suitable databases for you.

Hello, could you please describe the scope of your study?

We want to

Hello! Based on your interest in COVID-19 research, I recommend exploring these specific databases and public health:

- Tampere University Hospital
- UK Biobank
- Norwegian health registries at University of Oslo

These databases will provide you with valuable resources to support your study.

Are there any other concepts you'd like to add? If yes, please add them. If no, simply


Demographics Data Domains Data Provenance Observation Period Visit **Concept Browser** Meta Data

The concepts browser allows you to search for concepts by name or concept_id in all the data sources you select. No exact number of patients or occurrences are provided but they are rounded up to the nearest multiple of 100. Note that the number of patients and occurrences are for that concept only and does not include the children of that code. By clicking on the concept_id you can view additional information about the concept in Athena

Concept Browser

Show 10 entries

concept_id	concept_name
35894915	COVID-19 vaccine
37311060	Suspected COVID-19
37311061	COVID-19
704996	Patient meets COVID-19 laboratory diagnostic criteria
705076	Post-acute COVID-19
586526	SARS-CoV-2 (COVID-19) RNA [Presence] in Nasopharynx by NAA with probe detection
706163	SARS-CoV-2 (COVID-19) RNA [Presence] in Respiratory specimen by NAA with probe detection
706179	SARS-CoV-2 (COVID-19) Ab panel - Serum or Plasma by Immunoassay
706181	SARS-CoV-2 (COVID-19) IgG Ab [Presence] in Serum, Plasma or Blood by Rapid immunoassay
723473	SARS-CoV-2 (COVID-19) IgA Ab [Presence] in Serum or Plasma by Immunoassay



Concept set definition

710158	COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
756061	Asymptomatic COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35894915	COVID-19 vaccine	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
37310268	Suspected COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
703431	COVID-19 excluded	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The concept set was successfully defined. You can either download it or send it to your ATLAS instance.

How can I help you?

Concept set definition

710158	COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
756061	Asymptomatic COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35894915	COVID-19 vaccine	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
37310268	Suspected COVID-19	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
703431	COVID-19 excluded	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

[Submit](#)

Showing 1 to 5 of 5 entries

<input type="checkbox"/>	Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	<input checked="" type="checkbox"/> Exclude	<input type="checkbox"/> Descendants	<input checked="" type="checkbox"/> Mapped
<input checked="" type="checkbox"/>	37310268	1240761000000102	Suspected COVID-19	Observation	Non-Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	35894915	OMOP5042939	COVID-19 vaccine	Drug	Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	703431	1321101000000103	COVID-19 excluded	Observation	Non-Standard	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	710158	U07.100	COVID-19	Condition	Non-Standard	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>	756061	OMOP4873910	Asymptomatic COVID-19	Condition	Non-Standard	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Cohort definition

Ok, can you help me with the cohort definition?



Regarding the observation window, what is the minimum number of days needed before the continuous observation? You must choose from 0, 1, 7, 14, 21, 30, 60, 90, 120, 180, 365, 548, 730 or 1095.

7 days is ok




How many days after event index date? You must choose from 0, 1, 7, 14, 21, 30, 60, 90, 120, 180, 365, 548, 730 or 1095.

30



The cohort template is complete. You can either download it or send it to your ATLAS instance. Thank you!

 Download

 Send to ATLAS

How can I help you?





Cohort definition

Ok, can you help me with the cohort definition?

Regarding the observation window, what is the minimum number of days needed before the continuous observation? You must choose from 0, 1, 7, 14, 21, 30, 60, 90, 120, 180, 365, 548, 730 or 1095.

How many days after event index date? You must choose f

The cohort template is complete. You can either download

Download



How can I help you?

Definition | Concept Sets | Generation | Samples | Reporting | Export | Versions

Hippocrates costum cohort definition

Cohort Entry Events

Events having any of the following criteria:

a condition occurrence of Covid Demo - JRA

with continuous observation of at least 7 days before and 30 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events

Cohort definition

```
{
  "ConceptSets": "Are there any other concepts you'd like to add? If yes, please add them.
If no, simply respond with 'no'.",

  "PrimaryCriteria": {
    "ObservationWindow": {
      "PriorDays": "Regarding the observation window, what is the minimum number of days
needed before the continuous observation? You must choose from 0, 1, 7, 14, 21, 30, 60,
90, 120, 180, 365, 548, 730 or 1095.",
      "PostDays": "How many days after event index date? You must choose from 0, 1, 7,
14, 21, 30, 60, 90, 120, 180, 365, 548, 730 or 1095."
    }
  }
}
```



Cohort definition

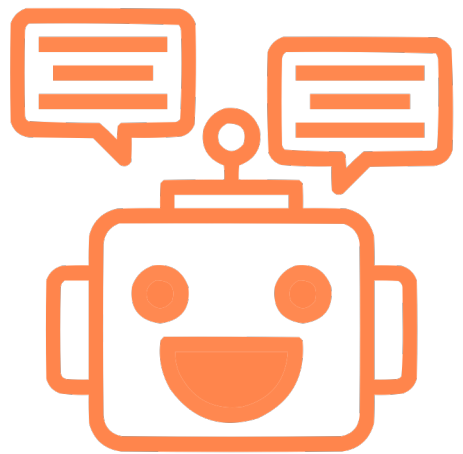
```
{
  "ConceptSets": "Are there any other concepts you'd like to add? If yes, please add them.
  If no, simply respond with 'no'.",

  "PrimaryCriteria": {
    "ObservationWindow": {
      "PriorDays": "Regarding the observation window, what is the minimum number of days
      needed before the continuous observation? You must choose from 0, 1, 7, 14, 21, 30, 60,
      90, 120, 180, 365, 548, 730 or 1095.",
      "PostDays": "How many days after event index date? You must choose from 0, 1, 7,
      14, 21, 30, 60, 90, 120, 180, 365, 548, 730 or 1095."
    }
  }
}
```

Processed by the LLM until it returns an acceptable answer



Ongoing work



Researchers interested in conduct a study