Unlocking Healthcare Insights: Building a Data Pipeline in Airtable

Solmaz Eradat Oskoui, PhD

Overview

1	Background & Company Mission
2	OHDSI Community & OMOP-CDM
3	The No-Code Landscape
4	Why Airtable?
5	Source to Standard Mapping
6	Mapping Approaches within Airtable
7	Overview of Data Pipeline
8	Live Demo
9	Conclusion

Background

Applied Mathematics, PhD

St.Andrews University

2014

Consultant/Senior Data Analyst/Data Engineer

Academia/IT/Banking/Healthcare

2014 - 2023

Founder

Data Harmonise Ltd

2024

Company Mission

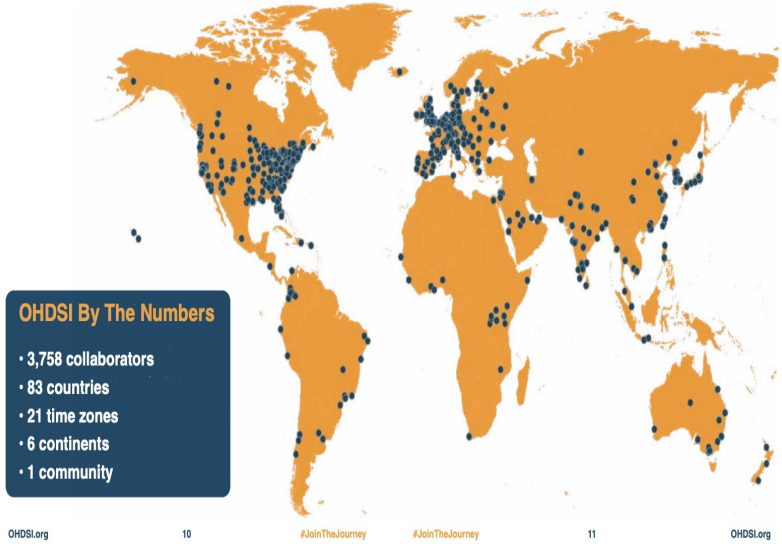
*Advance **interoperability** and **quality** within healthcare data by providing **data standardisation** and **engineering solutions** leveraging the Observational Medical Outcomes Partnership Common Data Model (**OMOP-CDM**). Our commitment is to facilitate trusted data links, foster collaboration, accelerate research outputs, and drive innovation.*

OHDSI Community & OMOP-CDM

- **Mission:** The Observational Health Data Sciences and Informatics (**OHDSI**) a multi-stakeholder, interdisciplinary **community** striving to improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

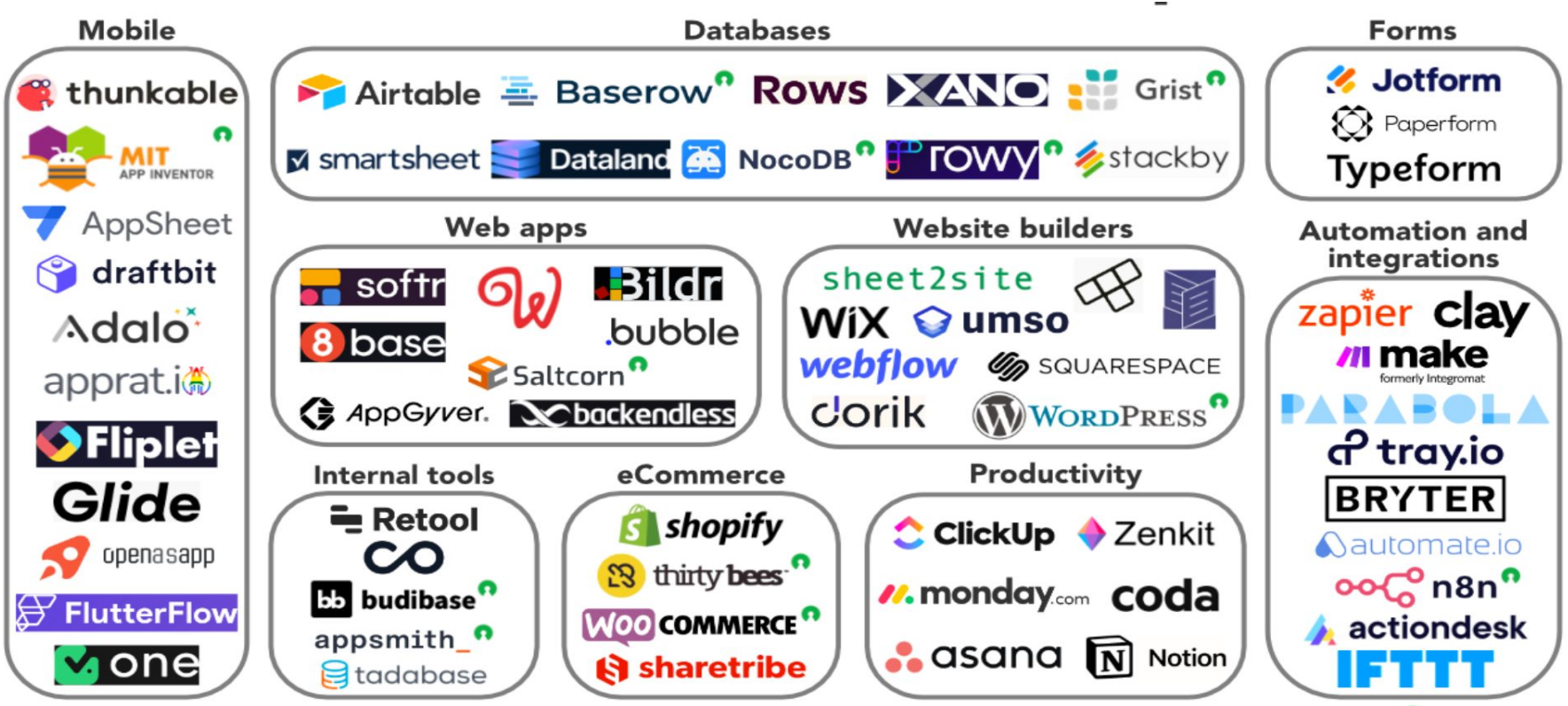
- **Method:** Creating a standardised data model the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM).

- **Impact:**
  - **Growing adoption over 10 years**
  - **534 data sources** from **49 countries mapped**
  - Represents **956 million patients** (**12% of world population)**



Source: [Where the OHDSI community has been and where are we going Annual report 2023]

The No-Code Landscape



Source: [The no-code landscape: 100+ platforms to try]

Why Airtable?

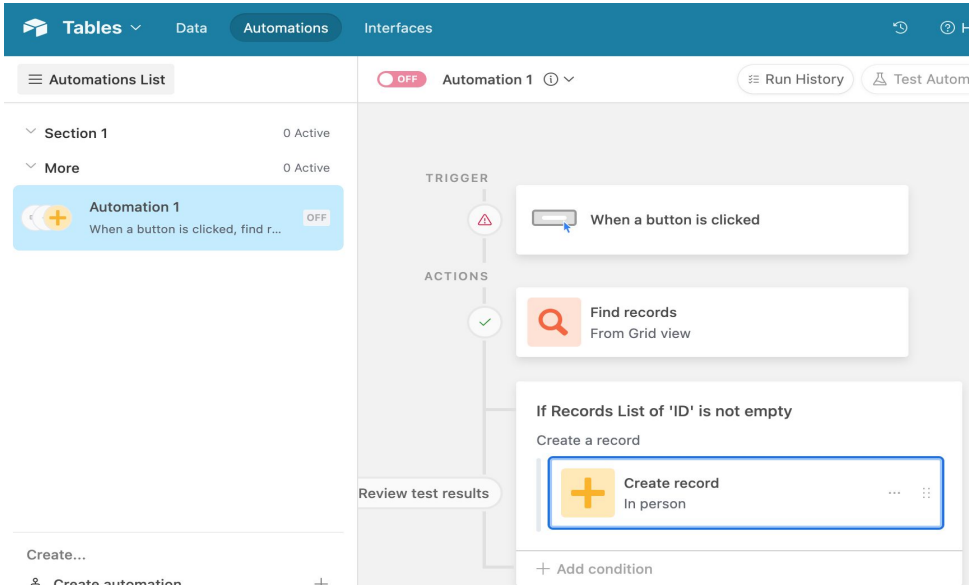
- Cloud-based collaboration platform with **low-code/no-code capabilities**.
- **Create workflows** for enhanced data organisation and management.
- Combines the **functionality of a spreadsheet with the features of a database**.

Within a workspace a user can add:

- **Base (collection of data tables)**
- **Interface (a visual representation of base data)**

Workflows can be automated using:

- **Actions (run scripts)**
- **Triggers**



Airtable

Free

For individuals or very small teams just getting started with Airtable

Free

✓ Current plan

Free includes:

- ✓ Unlimited bases
- ✓ 1,000 records per base
- ✓ Up to 5 editors
- ✓ 1 GB of attachments per base
- ✓ 100 automation runs
- ✓ Interface Designer

Team Try for free

For teams building apps to collaborate on shared workflows

$20 per seat/month billed annually

Get **$48** in annual savings

Choose Team

Everything in Free, plus:

- ✓ 50,000 records per base
- ✓ 25,000 automation runs
- ✓ 20 GB of attachments per base
- ✓ Standard sync integrations
- ✓ Extensions
- ✓ Gantt and timeline view
- ✓ Expanded color, formatting, and calendar options

Business

For teams and departments who need advanced features and basic administration

$45 per seat/month billed annually

Get **$108** in annual savings

Choose Business

or contact sales

Everything in Team, plus:

- ✓ 125,000 records per base
- ✓ 100,000 automation runs
- ✓ 100 GB of attachments per base
- ✓ Premium sync integrations
- ✓ Verified data
- ✓ Two-way sync
- ✓ Admin panel
- ✓ SAML-based single sign-on

Enterprise Scale

For organizations looking to build scalable, flexible apps with powerful governance and administration

Contact sales

Everything in Business, plus:

- ✓ 500,000 records per base
- ✓ 500,000 automation runs
- ✓ 1,000 GB of attachments per base
- ✓ On-premises sync integrations
- ✓ Enterprise Hub
- ✓ Enhanced security and admin controls
- ✓ Enterprise API
- ✓ Extension and integration management
- ✓ Audit logs and DLP

Source to Standard Mapping

Extract

Transform

Load

patients.csv

person.sql

person table

Mapping Approaches within Airtable

Airtable doesn't currently support directly creating new tables using SQL scripts.

Approaches	Pros	Cons
1. Airtable Import Utility: Run SQL script locally on source data and import CSV file	• No coding required in Airtable • Using Airtable as a data storage	• Requires latest data source to be downloaded locally, creating data redundancy • Addition of manual steps to maintain the data pipeline
2. Airtable Automation with Script Block: Setup automations triggered by specific events	• Visibility of transformation logic in Airtable. E.g. script block that uses the Airtable Formula language to manipulate data.	• Requires knowledge of airtable scripting language • Lack of integration of Git for version control supporting continuous integration and continuous development (CI/CD).
3. Airtable-API: Integrate data in Airtable with any external system	• Many use cases using Python libraries like **pandas** with **airtable-api** to List, Retrieve, Create, Update/Upsert and Delete records.	• Requires Python and API knowledge to interact with databases, transform data, and upload via Airtable API.

API - Application Programming Interface

Overview of Data Pipeline

Preparation:

- **Airtable Setup:**
  - **Airtable base** and **table** with the data you want to extract.
  - **Airtable API key** from your Account Settings.
- **Python Environment:**
  - Utilise the `requests` library to make API calls to the Airtable endpoint for desired table.

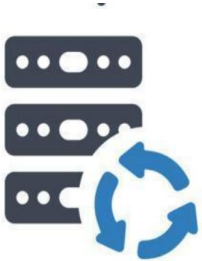
Data Extraction:

- **API Calls:**
  - Construct the URL (Uniform Resource Locator) using the Airtable **API base URL**, **base ID**, and **table ID**.
- **Pagination:**
  - Airtable limits records per request. Use the `offset` parameter in the URL to retrieve all records iteratively.
- **Data Parsing:**
  - Parse the JSON response from the API call to extract the list of records.Each record will be a dictionary with its fields and an ID.

Data Transformation:

- **Extract fields:**
  - Extract fields from records and transform to a dataframe.
- **Temporary databases:**
  - Load dataframe into temporary database in memory.
  - Run SQL script on temporary database and load transformed data into a placeholder database in memory.

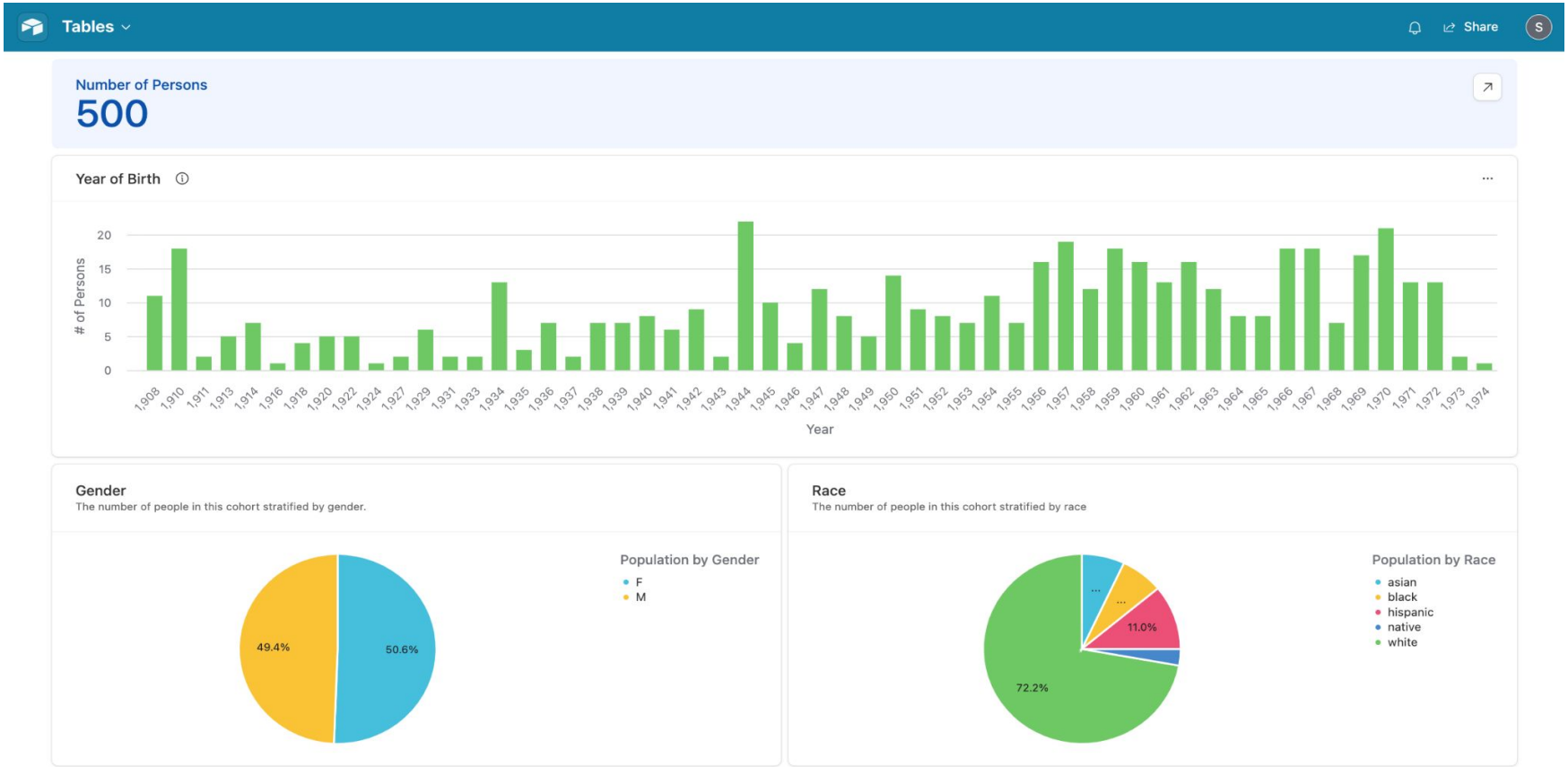
Overview of Data Pipeline ...



Data Load:

- **Airtable Setup:**
  - Placeholder **Airtable base** and **table** with the data you want to post.
  - **Airtable API key** from your Account Settings.
- **API Calls:**
  - Construct the URL using the Airtable **API base URL**, **base ID**, and **table ID**.
  - Use a post request to load records transformed database to the Airtable base.

Live Demo


Tables
Share
S
Number of Persons
500
Year of Birth
# of Persons
Year
Gender
The number of people in this cohort stratified by gender.
49.4%
50.6%
Population by Gender
F
M
Race
The number of people in this cohort stratified by race
11.0%
72.2%
Population by Race
asian
black
hispanic
native
white

Conclusion

Ease of Use:
Low-cost, low-code alternative for validating medical devices, healthcare applications, small healthcare datasets, and more to standards such as the OMOP-CDM.

Flexibility:
Airtable allows for **storing various data types** (text, numbers, attachments, etc.) in a flexible table structure. The data definition language (DDL) for the OMOP-CDM Person table required no manipulation.

Data Visualisation:
No-code interface builds dashboards that usually require complex coding, saving valuable user time. Compared to a tool like Atlas used to analyse OMOP-CDM data, it was very easy to build similar summary dashboard.

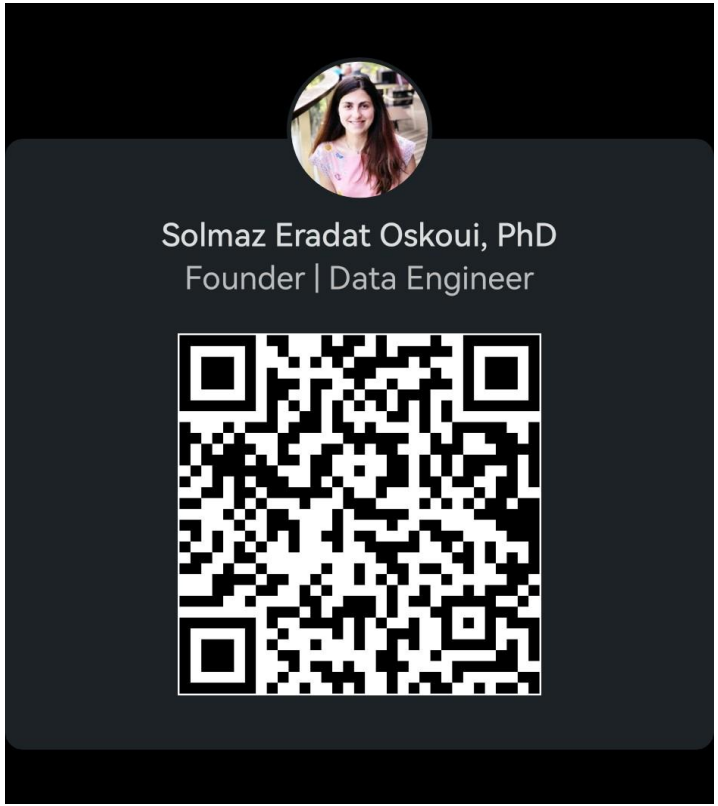
Limited Scalability:
Airtable might not be suitable for very large datasets or complex data pipelines.

Data Transformation Limitations:
Airtable offers basic data cleaning and filtering capabilities, but complex transformations might require additional tools.

Security Considerations:
Might not be as robust as dedicated data warehousing solutions for highly sensitive data.

GitHub Project

Thank you!



Email: solmaz.eradat@dataharmonise.co.uk

Useful links & Glossary

- ***Airtable Documentation***
- ***What is SQL?***
- ***What is an API?***
- ***Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM)***
- ***ATLAS***

- ***Workspace*** *- A collection of bases shared between a group of workspace collaborators.*
- ***Base*** *- A collection of data in Airtable, designed to contain all of the information a user needs for a project or workflow.*
- ***Base collaborator*** *- User which has access (at a specified permission level) to a specific base without necessarily having access to other bases within the same workspace*
- ***Interface*** *- An interface is a curated representation of base data created using Interface Designer. Interfaces are fully customizable and can contain various visual elements, data sources, and permissions.*
- ***SQL*** - Structured Query Language (SQL) is a programming language for storing and processing information in a relational database. In relational databases, data is stored in tables which are made up of columns and rows very similar to what is seen in Spreadsheets. Tables are made up of primary keys used as unique identifiers for each row of data. Rows from one table can be related to rows in a different table using a foreign key and are used to create a relationship between tables.
- ***API*** - Application Programming Interface (API) is a set of rules and protocols that enables software application to communicate with each other, to exchange data, features and functionality.