# Project Gropius

OMOP on Azure

## Exec Summary

OMOP is a common data model published by OHSDI, and currently very popular in the medical academic community. Although there is some adoption of OMOP it does not have the same industry wide support nor value of FHIR. I think everyone agrees that we should not change our strategy and Microsoft should continue to focus on FHIR. Given there are institutions that already have tooling or plan on making use of OMAP it makes sense we show customers how to create OMOP type solutions on Azure. We can build OSS best practice architectures and ARM templates which jumpstart a solution in Azure.

We are recommending the Smokejumpers build & publish an OSS architecture and ARM template on GitHub. This solution would leverage the Microsoft FHIR Server with Azure Data Factory connector to bulk export FHIR data into an OMOP schema in Azure Synapse and Azure Data Lake Storage. Optionally, a customer can use Azure Data Share to securely transport the data to other OMOP partners.

***Suggested stance*** *– Microsoft Health is building an OSS template for helping customers jumpstart FHIR & OMOP in Azure.*

## Background

Some of our research customers are looking for methods to standardize and share data in a consistent format. Through taking with several customers and accounts team the Observational Medical Outcomes Partnership (OMOP) Common Data Model from the Observational Health Data Sciences and Informatics (OHDSI, pronounced "Odyssey”) organization has gained popularity as the data model of choice in for some in academia.

Here is the reference link for more information on OHDSI and OMOP.

OHDSI - <https://www.ohdsi.org/>

OMOP - <https://www.ohdsi.org/data-standardization/>

OMOP Repo - <https://github.com/OHDSI/CommonDataModel>

#### Is OMOP a competitor to the Azure FHIR services offered by Microsoft?

No. OMOP is a data model for healthcare based primarily on Kimble Methodology. OMOP is not transactional like FHIR. FHIR holds richer data than OMOP.

#### Should Microsoft invest in OMOP the same way we did with FHIR?

No. OMOP is an open source schema. OMOP is not a fully adopted data standard like FHIR or HL7.

#### Who is using, building or evaluating OMOP?

The request for OMOP on Azure is coming primarily from our academic medical research institutions. However, some of the leading medical institutions with strong research areas are looking into OMOP as a method of sharing data with the academic institutions. Here are a few of the most recent customers asking about OMOP on Azure.

1. UCSF
2. University of Florida
3. Rush Medical Center
4. Cleveland Clinic
5. Medical University of South Carolina
6. University of Penn Medical
7. Johns Hopkins
8. Partners Health

#### Who is contributing to OMOP?

According to the OHDSI website there are several organizations listed as contributors to OHDSI. OMOP is one of many projects from OHDSI. OMOP is the biggest and most well-known. Here is the break down as of November 20th, 2019.

|  |  |  |
| --- | --- | --- |
| **Contributor Types** | **Count** | **% of Total** |
| Academic | 68 | 57.14% |
| ISV/SI/ Cloud | 20 | 16.81% |
| Pharma | 10 | 8.40% |
| Government Agency | 7 | 5.88% |
| Provider | 6 | 5.04% |
| Other | 5 | 4.20% |
| Non-Profit | 2 | 1.68% |
| Payor | 1 | 0.84% |
| **Grand Total** | **119** | **100.00%** |

This lines up with the type of customers asking for OMOP on Azure.

#### What are our competitors doing with OMOP?

AWS – is a contributing member to OHDSI. AWS has created a few blog posts with conceptual architectures for building OMOP on AWS.

Blog 1 - <https://aws.amazon.com/blogs/big-data/build-a-healthcare-data-warehouse-using-amazon-emr-amazon-redshift-aws-lambda-and-omop/>

Blog 2 - <https://aws.amazon.com/blogs/machine-learning/map-clinical-notes-to-the-omop-common-data-model-and-healthcare-ontologies-using-amazon-comprehend-medical/>

GCP – The only 2 data points for work by GCP on OMOP is:

1. The OMOP GitHub has a BigQuery DDL\*. The GitHub repository also has as DDL for Azure DW and SQL Server.
2. UCSF telling us they are ‘talking’ to our friends down the street about OMOP on their cloud. We can only assume UCSF was referring to GCP.

IBM – IBM has built OMOP support into their Unified Data Model for Healthcare. UDM is a product by IBM. IBM is also a contributor to OHDSI.

Oracle – Oracle is a contributor to OHSDI and has a DDL on the OMOP GitHub repo.

A note on contributors’ status – OHDSI lists everyone possible on the contributor page for OHDSI. However, I have confirmed with team members that US government entities like the VA have abandoned OMOP. Therefore, current active participation cannot be confirmed.

## Hypothesis

If we, Microsoft, builds a reference architecture and OSS ARM template for FHIR & OMOP on Azure we will

1. Unlock new healthcare data workloads on Azure
2. Create a method for OMOP partners to share data in a secure environment
3. Continue to win mindshare showing Microsoft is serious about Healthcare

## Suggested Solution

The suggested solution for OMOP on Azure is to leverage the Azure services we already have and provide our user community with multiple answers over the same data without requiring multi-master datasets. Create an OSS best practices architecture and template for use by customers. OMOP is one of the many industry common data models. Microsoft cannot productize each of them. We can create best practices, reference architectures and deployable OSS templates as we see demand.

An OMOP solution on Azure could consist of 5 basic elements.

1. Microsoft FHIR Server with Azure Data Factory connector
2. Azure Data Factory for mapping between the FHIR Server and other data source to the OMOP schema
3. Azure Synapse for hosting the OMOP schema and for converting the OMOP tables into flat files via Polybase.
4. Azure Data Lake Storage Gen 2 for storing the flat files.
5. Azure Data Share as an optional piece for sharing data with other OMOP partners securely.

OSS FHIR Server with ADF

Azure Data Factory

Azure Synapse

Azure Data Lake Storage

Azure Data Share

The Smokejumpers would build the ARM template, conceptual design, documentation and any other necessary assets. Customers would be responsible security, compliance, and updates.