

2023 Data Camp CCSQ D&A Roadmap & Vision Q&A

1. What governance policies will be in place to ensure data lineage is coming from a trusted source?

A - The first phase of this is tracking. Seeing the value of what transformation has occurred along the path of a particular file is one of the most critical elements. Now, we will be able to track that within the lineage capabilities that we are standing up. Generically, this is a continual improvement.

2. Data transfer from the Centralized Data Repository (CDR) to Integrated Data Repository (IDR) by touching the mainframe is a huge pain point and cumbersome. Would like CMS to make this less burdensome and more seamless.

A - The current model is to connect to an Analytic Data Mart (ADM) to your analytics over in the IDR and move your data out of the IDR.

This is the short-term model. We are working with the IDR team when they are using IDR in the cloud to connect into that instance, using Snowflake, etc. Improvements are scheduled for November.

We are also working with the IDR team to see if there is an interim solution we can do where we directly submit jobs to the IDR without Snowflake.

3. Will there be seamless integration between Databricks and QuickSight?

A - Yes. There is a recent plugin available for that, and there is also the ability for Databricks to write to a common location that that can access and update that way.

4. Why are the CDR and IDR both running a similar data system? Are there plans to combine them and keep the upsides of either of them or both?

A - The IDR has mature claims data and additional datasets, and the CDR team also has our National Claims History (NCH) feed and our claims data that we keep within our CDR. The long-term goal is to align with CMS' approach to claims. Once CMS designates the agency's claims repository, we will connect to it and deprecate our claims repository. The NCH feed, Medicare Quality Assurance (MQA) feed and Medicare Quality Release (MQR) feed are all potentially going away. Currently, there is no road map or timetable for this removal.

5. Will IDR Part C data be added to the CDR at some point?

A - Both will be made available. We are currently working with the necessary agency, representatives on getting our data added to the CDR. Whether that is in the form of the IDR or a separate file feed like the current NCH feed is to be determined. Users will need a Data Use Agreement (DUA) that says they are allowed to see Parts C claims regardless of either approach.

6. Will the data catalog documentation be improved in the near future?

A - This is an active project. Right now, each line of business and data owner produces their own data catalog which leads to discrepancies. We are currently in the process of creating standardized documentation. We do not want to design this in a vacuum so let us know if you are invested in it, sign up and we would love to talk to users about it in the future.

7. Will National Healthcare Safety Network (NHSN) data be added to the CDR at some point?

A - This data is currently in the CDR. It is a manually uploaded from the Centers for Disease Control (CDC).

8. Is there some plan to allow users access to the Databricks File System (DBFS) or will there be some process to expedite mounting of Simple Storage Service (S3) buckets to Databricks?

DBFS is not enabled as it does not have anti-virus (AV) capabilities. To expedite mounting S3 buckets to Databricks users must fill out a Decision Analysis and Resolution (DAR) request.

We are looking for macros to make it easier for users to look up their bucket path so that they can use it in Databricks a lot easier.