

OSP – OHIP Number Data Quality Issue

June 14, 2023

On April 20th, following the FY 22/23 data submission, DATIS discovered duplicate client records were being created when NLO data submissions were integrated with historical data in the OSP database. Investigation into the issue revealed large-scale data quality issues with client OHIP numbers being submitted.

Client OHIP numbers are a very sensitive data element as they are the only unique client identifier in the OSP database. For this reason, the OHIP_NO field is never overwritten in the OSP database. To complicate matters, having an OHIP number is not a requirement for service; therefore not mandatory in the OSP database. Per the OSP Data Dictionary (V1.13), data entry guidance for client records with no OHIP number is to submit a unique Client ID. Further investigation has revealed that a large number of records submitted to the OSP database do not contain a truly unique Client ID. The Client ID field is critical when there is no OHIP number submitted.

Data ingestion process

All client records submitted to the OSP database go through an ETL (extract, transform, and load) process prior to being ingested into the database. The ETL uses pre-programmed rules to clean and organize data before it is loaded into the OSP database, such as checking each incoming record for an org_id (organization ID). Another function of the ETL is assigning all clients an auto-generated, unique Client Key that ties all records for a client together, and is used in reporting. Client Keys are generated using the client's OHIP number. In the absence of an OHIP number, the logic then looks to use the Client ID. If both the OHIP number and Client ID are absent, the Row ID for that particular incoming record will be used. DATIS recognizes the limitations associated with generating a Client Key from the Row ID that will change with each submission.

Analysis of OHIP number and Client ID

Extensive analysis of client OHIP numbers and Client IDs submitted to the OSP database was conducted by DATIS and as a result, uncovered the following problematic scenarios:

Scenario number	Scenario	Result	Impact to the data
1	OHIP_NO and/or Client ID submitted with value = 0, n/a, N/A Ex. OHIP_NO = 0; Client ID = n/a	OHIP_NO submitted with a non-unique value will generate a non-unique Client Key. For example, if records for 5 different clients are submitted with OHIP_NO = 0, the same Client Key will be generated for each client, and the records for those 5 different clients will be tied together as one client. Including a non-unique Client ID = n/a for each of the 5 clients indicates they are all the same client.	Multiple clients counted as one client, resulting in an under-representation of clients – reports results will be lower than actual.

2	OHIP_NO submitted as null and subsequently updated in a later submission with OHIP_NO (or vice versa).	The Client Key generated for the incoming record with no OHIP_NO (based on Client ID or Row ID) does not match the original Client Key generated and results in a 'new' (duplicate) client.	One client counted as multiple clients, resulting in an over-representation of clients – report results will be higher than actual.
3	OHIP_NO submitted in original submission is updated with spaces, letters, and/or special characters. Ex. Original submission = 1234 567 890, subsequent submission includes version code = 1234 567 890 RG or valid OHIP_NO per the data specs = 1234567890.	The Client Key generated for the incoming record (based on OHIP number) does not match the original Client Key generated and results in a 'new' (duplicate) client.	One client counted as multiple clients, resulting in an over-representation of clients – report results will be higher than actual.
4	Two or more OHIP_NO are submitted for the same client. This could be a result of transposing two of the numbers when entering.	The Client Key generated for the incoming record (based on OHIP_NO) does not match the original Client Key generated and results in a 'new' (duplicate) client.	One client counted as multiple clients, resulting in an over-representation of clients – report results will be higher than actual.
5	OHIP_NO submitted is invalid per the OSP Data Dictionary.	Invalid OHIP_NO in the OSP database. If this invalid OHIP_NO is consistently submitted, there is no implication to the data.	Providing the invalid OHIP_NO is consistently submitted, there will be no impact on reports. This remains a DQ issue.

Removing duplicate records

It is critical that the duplicate records in the OSP database are removed prior to the next data submission. DATIS has built automated processes and performed the following actions for each of the scenarios named above:

Scenario number	Scenario	Action/Result
1	OHIP_NO and/or Client ID submitted with values = 0, n/a, N/A	Update all OHIP_NO and/or Client ID = 0, n/a, N/A to null (blank) to regenerate a unique Client Key for each record based on client first name, last name, date of birth.
2	OHIP_NO submitted as null and subsequently updated in a later submission with OHIP_NO (or vice versa)	Merge duplicate client records where Client IDs match; or client first name, last name, date of birth, and referral received date matches.
3	OHIP_NO submitted with in original submission is updated with spaces, letters, and/or special characters.	Remove all spaces, letters, and/or special characters, and in cases where more than 10 digits are left, remove any digits after

	Ex. original submission = 1234 567 890, subsequent submission includes version code = 1234 567 890 RG	the first 10. Regenerate the Client Key based on transformed OHIP_NO. Merge all records with matching OHIP_NO.
4	Two or more OHIP_NO are submitted for the same client. This could be the result of a typo such as transposing two of the numbers when entering.	Merge duplicate client records where Client IDs match; or in cases where one of the OHIP_NO differs by one digit and client first name, last name, date of birth match. OHIP_NO most recently submitted is kept. Cases with more variance between records are not merged and require review by NLO.
5	OHIP_NO submitted is invalid per the OSP Data Dictionary.	Remove all spaces, letters, and/or special characters, and in cases where more than 10 digits are left, remove any digits after the first 10. OHIP_NO that are less or more than 10 digits will remain in the OSP database.

Where applicable, a package will be sent to each NLO with details on the actions performed.

The resolution of this DQ issue will require both an interim and permanent solution. The end goals are:

- All OHIP numbers in the OSP database are per the data specs – 10 digits.
- All client records are uniquely identified to allow cross validation of incoming records when no or different OHIP numbers are submitted.

Interim solution

The interim solutions named below will apply to the June 2023 data submission, and all subsequent data submissions until a permanent solution is in place.

To mitigate the creation of duplicate client records when data is integrated into the OSP database, DATIS has implemented the following:

- Manual and automated check processes of all incoming data files prior to loading the data into the OSP database.
- New ETL logic programmed to:
 - Remove all spaces, letters, and/or special characters from incoming OHIP_NO
 - Reduce all OHIP_NO greater than 10 digits to 10 digits (keeping only the first 10 digits)
 - Transforming OHIP_NO submitted as 0, n/a, N/A to null (blank)

DATIS is requesting all NLOs/LI service providers to take the following actions:

- Update (if possible) all invalid or incorrect OHIP_NO – **please provide the updated OHIP numbers to DATIS prior to the next data submission to ensure the OSP data is updated prior to receiving your data.**
- Leave all invalid or incorrect OHIP_NO that you are unable to update as they are (do not make any changes) to ensure the records in your data submission match the records in the OSP database.

Going forward, with the updated ETL logic, all OHIP numbers with spaces, letters, and/or special characters, and OHIP numbers greater than 10 digits will be accepted and transformed to align with the data specs. This will eliminate any manual intervention needed to correct OHIP numbers prior to submitting data.

Permanent solution

At this time, a permanent solution has not yet been put in place. DATIS will work to identify an effective solution in consultation with the MHA Centre of Excellence. Please stay tuned for further communications.