Redivis: A collaborative, inter-institutional data commons

Ian Mathews
co-founder • ian@redivis.com
Vision

We imagine a world where policy and scientifically relevant data are readily available and interpretable for a wide range of individuals.

- Promote and facilitate collaboration
- Improve communication of data science
- Reduce fundamental tension between data availability and security
A five year journey...

- Began with data visualization
- How can we tell the stories within data?
- The problem wasn’t generalizable.

What we were building presupposed a foundation that didn’t exist.
Discovery
+ Metadata
+ Documentation
Administration

+ Attribution

+ Access Control
Collaboration  
+ Reproducibility  
+ Version control  
+ Interoperability
... and so much more
Each is a monumental challenge in its own right.
So we built on the foundation that already existed.
• Dataverse, CKAN, Dryad, Kaggle
• Metadata and ontologies: schema.org, SNOMED, OMOP
• Scalable access: attribute based access control (ABAC)
• Google Cloud. BigQuery.
• Kubernetes. Containers. Git.
• R, Python; Jupyter
• REST APIs
• The modern web. React.
• UX and UI design!
And after (a bit) of trial and error
Developed a platform that:

**Lets organizations:**
- Host data — from kilobytes to petabytes
- Public to high-risk PHI
- Handle access and security at scale
- Version and archive data

**Empowers researchers:**
- Discover and assess new data
- Query and merge terabytes of data in seconds
- Collaborate in real time
- Interface via R, Python, Jupyter, SQL
  - And, SAS / STATA / SPSS
(brief)

DEMO
<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Description</th>
<th>Rows</th>
<th>Columns</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>MarketScan Inpatient Admissions</td>
<td>Inpatient admission summary records. Inpatient meaning that the patient is fo..</td>
<td>1 (1)</td>
<td>191</td>
<td>8 (1)</td>
</tr>
<tr>
<td>Optum SES Inpatient Confinement</td>
<td>Member inpatient confinements</td>
<td>1 (1)</td>
<td>78</td>
<td>5 (1)</td>
</tr>
<tr>
<td>STARR Tahoe Procedures</td>
<td>Coded procedures from SHC and SCH both from Epic and prior EMRs. This ta..</td>
<td>1 (1)</td>
<td>68</td>
<td>17 (10)</td>
</tr>
<tr>
<td>MarketScan Medicare Supplemental Insurance</td>
<td>Inpatient admission summary records. Inpatient meaning that the patient is fo..</td>
<td>1 (1)</td>
<td>74</td>
<td>0</td>
</tr>
<tr>
<td>MarketScan Inpatient Services</td>
<td>This file contains individual inpatient claim records that were used to create a..</td>
<td>1 (1)</td>
<td>206</td>
<td>7 (1)</td>
</tr>
<tr>
<td>AFC Patient Procedure</td>
<td>Placeholder</td>
<td>1 (1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Optum Procedure Lookup</td>
<td>Procedure Lookup Table - Descriptions and Categories for Procedure Codes</td>
<td>1 (1)</td>
<td>312,784</td>
<td>2 (8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (1)</td>
<td>211</td>
<td>6 (6)</td>
</tr>
</tbody>
</table>

#OpenScienceinAction | www.theneuro.ca
Optum SES Inpatient Confinement

Overview

Introduction

Member inpatient confinements

Data

Download (limited)

Documentation

Methodology

The Inpatient Confinement table contains a summarized record for each inpatient episode occurring in an acute care hospitalization or skilled nursing facility setting. There is a unique record for every hospitalization observed during the data period. In this unique confinement record, all facility detail records are bundled and reported in a single unduplicated row. Included in the record are the admission date (ADMITDATE), discharge date (DISCHDATE), length of stay (LOS), primary diagnosis (DIAG1), provider ID (PROV), place of service (POS), discharge status (DSTATUS), and standardized cost (STD_COST).

There are up to five ICD-9 diagnosis codes entered into the Inpatient Confinement table (DIAG1-DIAG5). The variable DIAG1 is typically considered the primary diagnosis.

By using the confinement ID (CONF_ID), it is possible to obtain claims for those professional services that have been rendered and billed separately during the confinement period from the Medical Claims table. In a limited number of cases, no records will be found in the medical table to correspond with the confinement. Sometimes a medical claim will be denied after having been paid and included in the CDM long enough for the confinement to be built. When this happens, the medical claims are no longer be included in Data Mart, but the confinement ID still exists.

If the patient is transferred to a different facility or different acuity setting within the
<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conf_Id</td>
<td>string</td>
<td>Confinement Id</td>
</tr>
<tr>
<td>Copay</td>
<td>float</td>
<td>Copay Amount</td>
</tr>
<tr>
<td>Deduct</td>
<td>float</td>
<td>Deductible Amount</td>
</tr>
<tr>
<td>Diag1</td>
<td>string</td>
<td>Diagnosis 1</td>
</tr>
<tr>
<td>Diag2</td>
<td>string</td>
<td>Diagnosis 2</td>
</tr>
<tr>
<td>Diag3</td>
<td>string</td>
<td>Diagnosis 3</td>
</tr>
<tr>
<td>Diag4</td>
<td>string</td>
<td>Diagnosis 4</td>
</tr>
<tr>
<td>Diag5</td>
<td>string</td>
<td>Diagnosis 5</td>
</tr>
<tr>
<td>Disch_Date</td>
<td>date</td>
<td>Discharge Date</td>
</tr>
<tr>
<td>Drg</td>
<td>string</td>
<td>Diagnosis Related Group</td>
</tr>
<tr>
<td>Dstatus</td>
<td>string</td>
<td>Discharge Status</td>
</tr>
<tr>
<td>Icd_Flag</td>
<td>integer</td>
<td>ICD Flag</td>
</tr>
<tr>
<td>Lpstatus</td>
<td>string</td>
<td>IP Status</td>
</tr>
<tr>
<td>Los</td>
<td>integer</td>
<td>Length of Stay</td>
</tr>
<tr>
<td>Pos</td>
<td>integer</td>
<td>Place of Service</td>
</tr>
<tr>
<td>Proc1</td>
<td>string</td>
<td>Procedure 1</td>
</tr>
<tr>
<td>Proc2</td>
<td>string</td>
<td>Procedure 2</td>
</tr>
<tr>
<td>Proc3</td>
<td>string</td>
<td>Procedure 3</td>
</tr>
<tr>
<td>Proc4</td>
<td>string</td>
<td>Procedure 4</td>
</tr>
</tbody>
</table>

If the patient is transferred to a different facility or different acuity setting within the...
## Optum SES Inpatient Confinement

### Overview

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PatId</td>
<td>integer</td>
<td>Patient Id</td>
</tr>
<tr>
<td>Pat_PlanId</td>
<td>integer</td>
<td>Patient Plan Id</td>
</tr>
<tr>
<td>Admit_Date</td>
<td>date</td>
<td>Admission Date</td>
</tr>
<tr>
<td>Charge</td>
<td>float</td>
<td>Charge Amount</td>
</tr>
<tr>
<td>Coins</td>
<td>float</td>
<td>Coinsurance Amount</td>
</tr>
<tr>
<td>Conf_Id</td>
<td>string</td>
<td>Confinement Identifier</td>
</tr>
<tr>
<td>Copay</td>
<td>float</td>
<td>Copay Amount</td>
</tr>
<tr>
<td>Deduct</td>
<td>float</td>
<td>Deductible Amount</td>
</tr>
<tr>
<td>Diag1</td>
<td>string</td>
<td>Diagnosis 1</td>
</tr>
<tr>
<td>Diag2</td>
<td>string</td>
<td>Diagnosis 2</td>
</tr>
<tr>
<td>Diag3</td>
<td>string</td>
<td>Diagnosis 3</td>
</tr>
<tr>
<td>Diag4</td>
<td>string</td>
<td>Diagnosis 4</td>
</tr>
<tr>
<td>Diag5</td>
<td>string</td>
<td>Diagnosis 5</td>
</tr>
<tr>
<td>Disch_Date</td>
<td>date</td>
<td>Discharge Date</td>
</tr>
<tr>
<td>Drg</td>
<td>string</td>
<td>Diagnosis Related Group</td>
</tr>
<tr>
<td>Dstatus</td>
<td>string</td>
<td>Discharge Status</td>
</tr>
<tr>
<td>Icd_Flag</td>
<td>integer</td>
<td>ICD Flag</td>
</tr>
<tr>
<td>Istatus</td>
<td>string</td>
<td>IP Status</td>
</tr>
</tbody>
</table>

### Data

Sort by: Index

Add to project
- Download (limited)

Optum
- Metadata access
  - Apply for access

My usage
- Studies: 2
- Projects: 2

© 2019 Redivis Inc.  Privacy  Terms

#OpenScienceinAction | www.theneuro.ca
```sql
SELECT t0.Patid AS Patid,
t0.Depression_date AS Depression_date,
t0.Delivery_date AS Delivery_date,
DATE_ADD(t0.Delivery_date, INTERVAL 6 month) AS delivery_6yr
FROM
    `Join depression & delivery output` AS t0
WHERE
    t0.Depression_date > t0.Delivery_date
    AND
    t0.Depression_date < DATE_ADD(t0.Delivery_date, INTERVAL 6 month)
```
Redivis: A collaborative, inter-institutional data commons

Ian Mathews
co-founder • ian@redivis.com