DPP Data Curation Pipeline Summary

# Domain 0 – Stored Functions

1. *A-CREATE\_analyst\_convert\_to\_AmDenv.sql*

**Inputs from analyst schema:** **NONE**

**Outputs:** CREATE FUNCTION *analyst.convert\_to\_AmDenv*

# Domain 1 – EHR Data

1. *B-Loading\_EHR\_data\_into\_SQL\_database.R*

**Inputs from analyst schema: NONE**

**Outputs:**

* *analyst.raw\_EHR\_demographics*
* *analyst.raw\_EHR\_race*
* *analyst.raw\_EHR\_encounters*
* *analyst.raw\_EHR\_pt\_prob\_list*

1. *C-CREATE\_analyst\_crosswalk\_uid\_to\_MRN.sql*

**Inputs from analyst schema:** *analyst.raw\_EHR\_demographics*

**Outputs:** CREATE TABLE *analyst.C\_crosswalk\_uid\_to\_MRN*

1. *D-CREATE\_analyst\_EHR\_Race\_tall.sql*

**Inputs from analyst schema:**

* + - *analyst.raw\_EHR\_race*
    - *analyst.C\_crosswalk\_uid\_to\_MRN*

**Outputs:** CREATE TABLE *analyst.D\_EHR\_race\_tall*

1. *E-CREATE\_analyst\_EHR\_demographics.sql*

**Inputs from analyst schema:**

* *analyst.C\_crosswalk\_uid\_to\_MRN*
* *analyst.D\_EHR\_race\_tall*
* *analyst.raw\_EHR\_demographics*

**Outputs:** CREATE TABLE *analyst.E\_EHR\_demographics*

1. *F-CREATE\_analyst\_EHR\_encounters.sql*

**Inputs from analyst schema:**

* *analyst.C\_crosswalk\_uid\_to\_MRN*
* *analyst.raw\_EHR\_encounters*

**Outputs:** CREATE TABLE *analyst.F\_EHR\_encounters*

1. *G-CREATE\_EHR\_pt\_prob\_list.sql*

**Inputs from analyst schema:**

* *analyst.C\_crosswalk\_uid\_to\_MRN*
* *analyst.raw\_EHR\_ pt\_prob\_list*

**Outputs:** CREATE TABLE *analyst.G\_EHR\_pt\_prob\_list*

# Domain 2 – Workflow Data

1. *H-CREATE\_analyst\_workflow\_data\_tall.sql*

**Inputs from analyst schema:** **NONE**

**Outputs:** CREATE TABLE *analyst.H\_workflow\_data\_tall*

1. *I-CREATE\_analyst\_textmessages\_tall.sql*

**Inputs from analyst schema: NONE**

**Outputs:** CREATE TABLE *analyst.I\_textmessages\_tall*

1. *J-CREATE\_analyst\_tm\_responses\_wide.sql*

**Inputs from analyst schema:** *analyst.I\_textmessages\_tall*

**Outputs:** CREATE TABLE *analyst.J\_tm\_responses\_wide*

1. *K-CREATE\_analyst\_tm\_response\_summarized.sql*

**Inputs from analyst schema:** *analyst.J\_tm\_responses\_wide*

**Outputs:** CREATE TABLE *analyst.K\_tm\_response\_summarized*

1. *L-CREATE\_analyst\_maps\_calls\_tall.sql*

**Inputs from analyst schema:** *analyst.C\_crosswalk\_uid\_to\_MRN*

**Outputs:** CREATE TABLE *analyst.L\_mapsCalls\_tall*

1. *M-CREATE\_analyst\_mapsCalls\_outcomes*

**Inputs from analyst schema:** *analyst.L\_mapsCalls\_tall*

**Outputs:** CREATE TABLE *analyst.M\_mapsCalls\_outcomes*

1. *N-CREATE\_analyst\_wide\_outcomes\_4\_analysis.sql*

**Inputs from analyst schema:**

* *analyst.H\_workflow\_data\_tall*
* *analyst.K\_tm\_response\_summarized*
* *analyst.C\_crosswalk\_uid\_to\_patient\_cd*
* *analyst.M\_mapsCalls\_outcomes*

**Outputs:** CREATE TABLE *analyst.N\_wide\_outcomes\_4\_analysis*

# Domain 3 – IncentaHealth Data

TBD

# Domain 4 – All Data

1. O-get\_curated\_data\_from\_db.R
2. P-create\_codebook.R
3. Q-Output curated datasets (split between for DM’s and for analysis – possibly an additional split for the level of PHI redaction)