

Run, Process, Append

Teamwork and Organization in State Reporting

Presented by: Amanda Merchant, AbdulHai Mohammed, and Liza Farrell



Outcomes

- Apply data retrieval strategies and methods
- Discuss best practices for state reporting processes and certified data warehousing
- Engage on or offline with others to learn how you all run, process and append your state reports.

Overview

Institution

“The Island University”

Founded 1947

Enrollment: 10,820

Faculty: 513

Staff: 888

Context

Terms: Traditional and Accelerated

Parts-of-Term: Several.

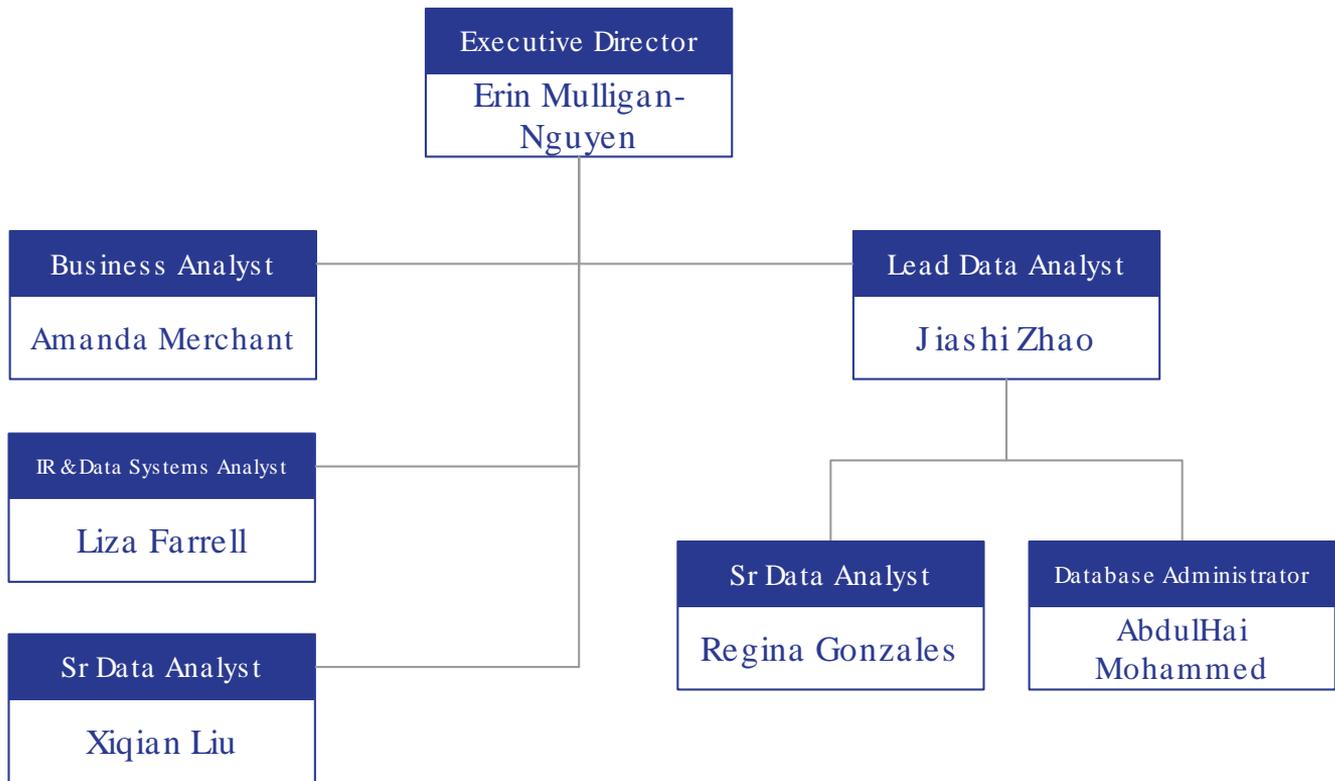
Flex: Got that too.

Reports

Reports we run for beginning of term processing:

- CBM00S
- CBM001
- CBM004

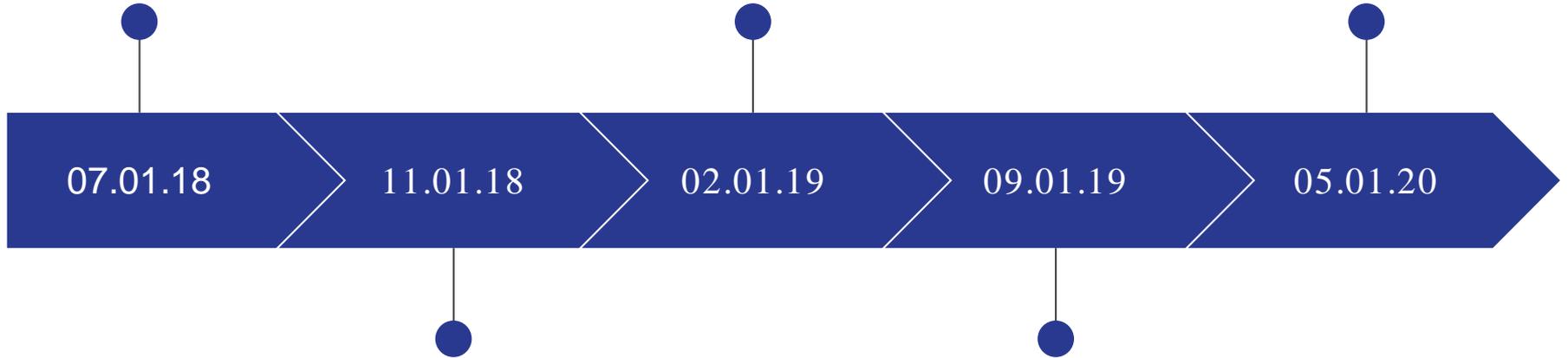
The team



We are 3.

Third analyst hired.
Now we are 5!

DBA hired. Now we
are 7 FTE!!!



Second analyst hired.
Now we are 4.

Fourth analyst hired.
Now we are 6!!

Challenges

Challenge 1

Run

Data Analyst spent a significant amount of time running the reports due to the number and we needed to document what we were doing and why.

Challenge 2

Process

Previous teams hadn't communicated with departments on campus when data issues arose. Lots of errors back on the reports, lots of time spent fixing them.

Challenge 3

Append

Previous teams had used Oracle to store reports and had kept only certain parts of the reports. We were not fully utilizing the data in analytics or research.

Run

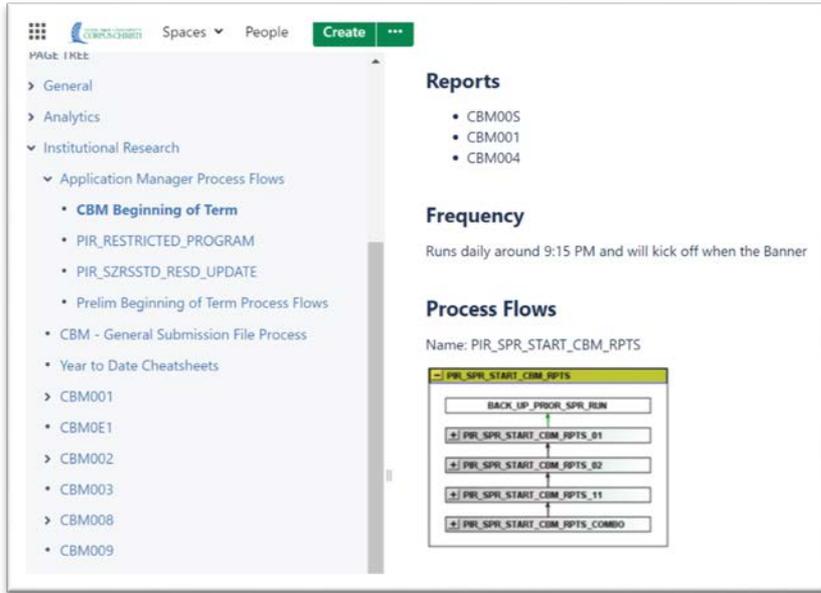
Addressing Challenge 1

- Documentation
- Handoff



Documentation - Confluence

App Manager Documentation



Navigation: Spaces, People, Create, ...

PAGE TREE

- General
- Analytics
- Institutional Research
 - Application Manager Process Flows
 - CBM Beginning of Term
 - PIR_RESTRICTED_PROGRAM
 - PIR_SZRSTDD_RESD_UPDATE
 - Prelim Beginning of Term Process Flows
 - CBM - General Submission File Process
 - Year to Date Cheatsheets
- CBM001
- CBM0E1
- CBM002
- CBM003
- CBM008
- CBM009

Reports

- CBM005
- CBM001
- CBM004

Frequency

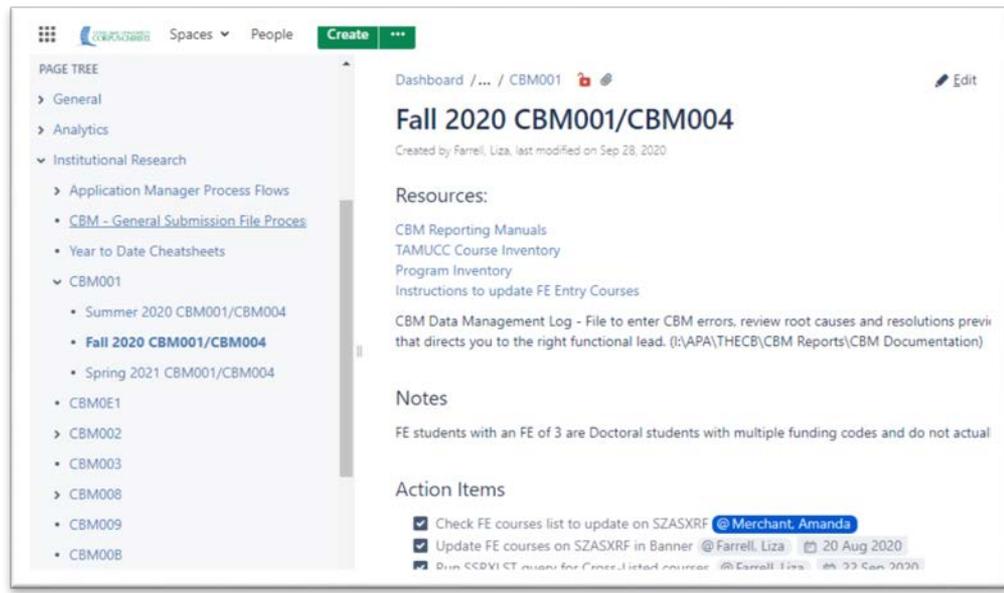
Runs daily around 9:15 PM and will kick off when the Banner

Process Flows

Name: PIR_SPR_START_CBM_RPTS

```
graph TD; A[BACK_UP_PRIOR_SPR_RUN] --> B[PIR_SPR_START_CBM_RPTS_01]; B --> C[PIR_SPR_START_CBM_RPTS_02]; C --> D[PIR_SPR_START_CBM_RPTS_11]; D --> E[PIR_SPR_START_CBM_RPTS_COMBO]
```

CBM Report Notes and Checklists



Navigation: Spaces, People, Create, ...

Dashboard / ... / CBM001

Fall 2020 CBM001/CBM004

Created by Farrell, Liza, last modified on Sep 28, 2020

Resources:

- CBM Reporting Manuals
- TAMUCC Course Inventory
- Program Inventory
- Instructions to update FE Entry Courses

CBM Data Management Log - File to enter CBM errors, review root causes and resolutions prior that directs you to the right functional lead. (APA/THC/CBM Reports/CBM Documentation)

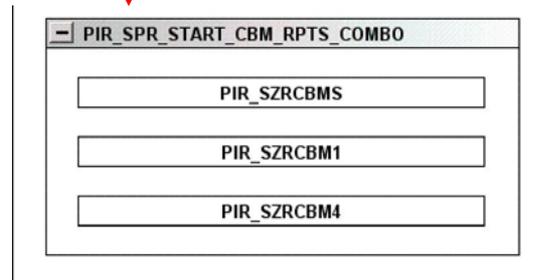
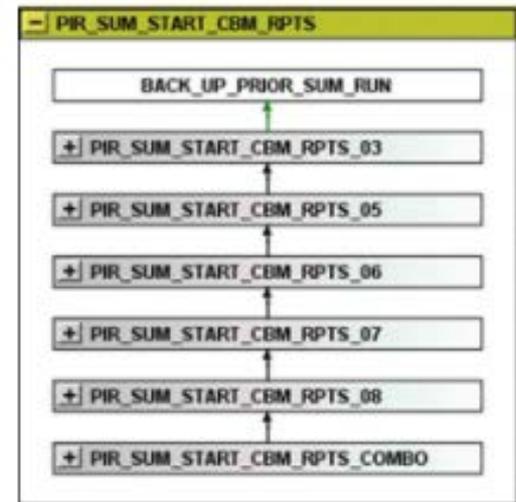
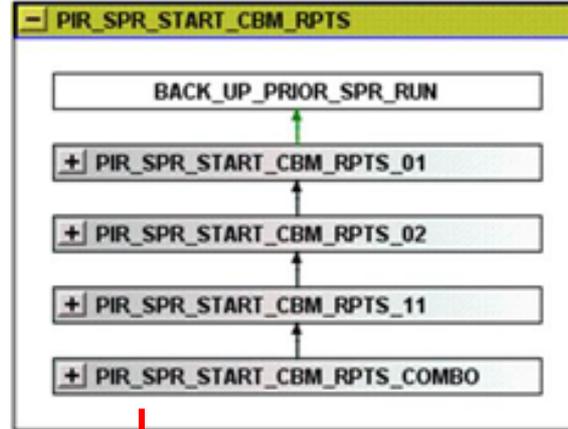
Notes

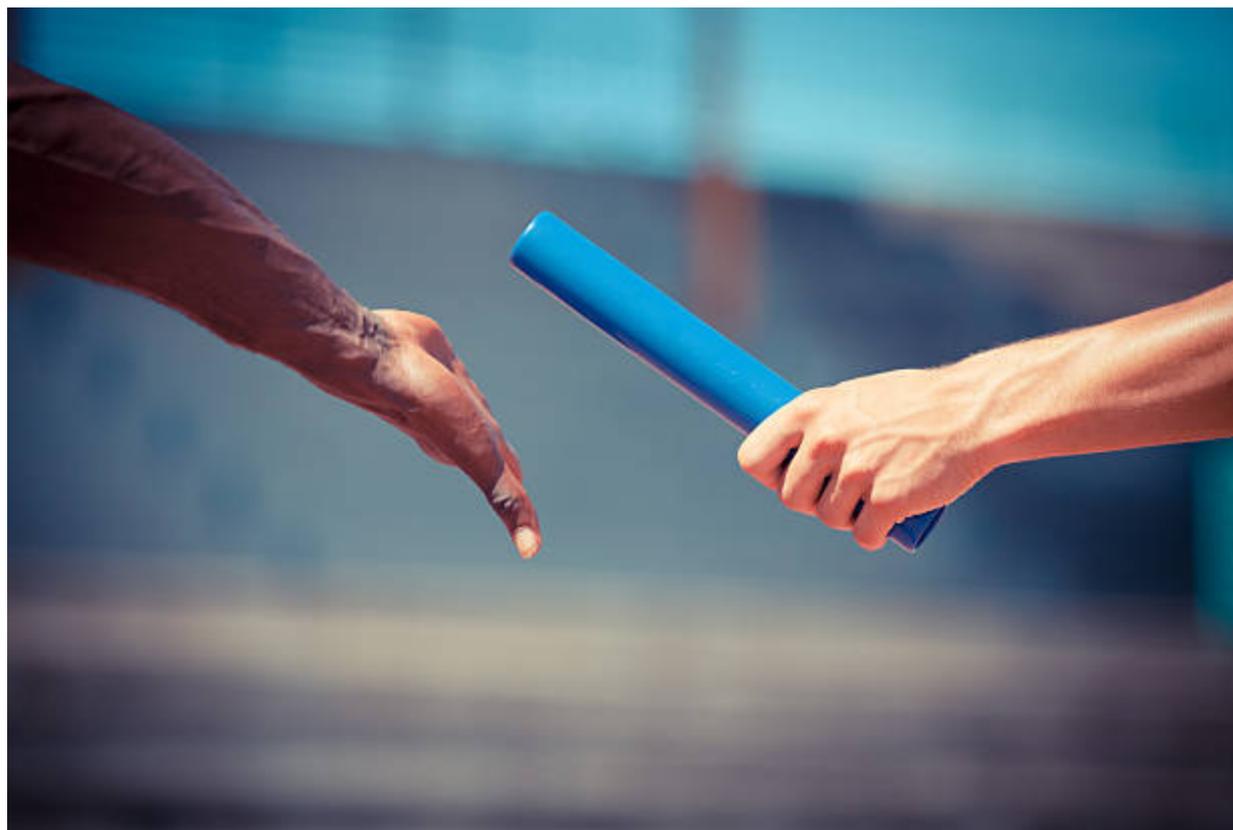
FE students with an FE of 3 are Doctoral students with multiple funding codes and do not actual

Action Items

- Check FE courses list to update on SZASXRF @ Merchant, Amanda
- Update FE courses on SZASXRF in Banner @ Farrell, Liza 20 Aug 2020
- Print CCQVICT menu for Counselor listed courses @ Farrell, Liza 20 Aug 2020

Outcomes - Application Manager Flows





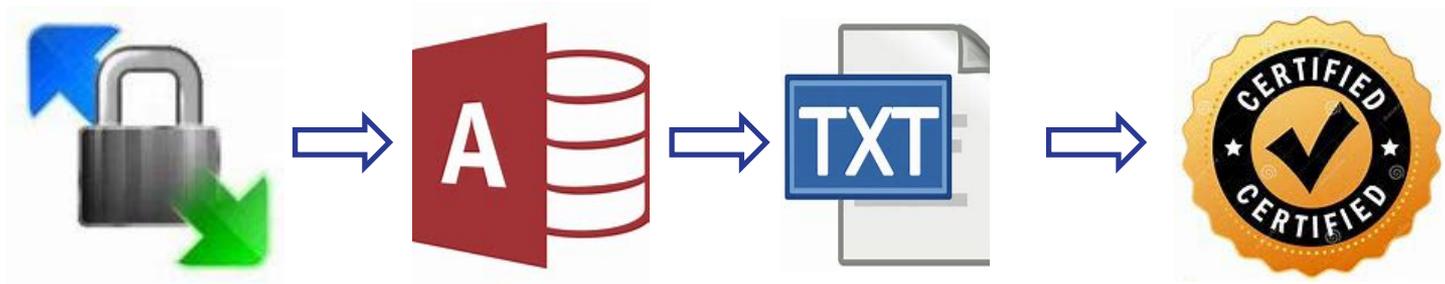
Process

Addressing Challenge 2

From .csv to .xlsx to .txt to Certified

Process – From .cvs to .xlsx to .txt to Certified

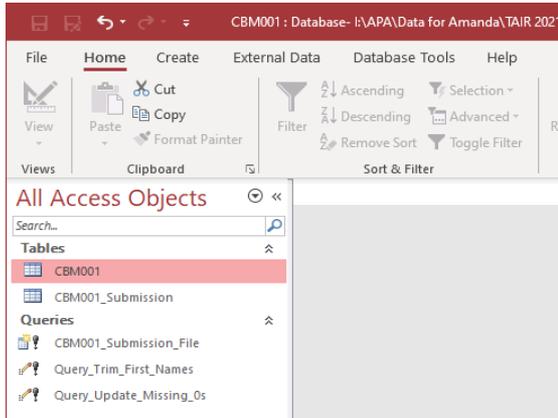
(For this presentation I will focus on the CBM001 Report for the 20th Class Day Census)



Process – Microsoft Access

(For this presentation I will focus on the CBM001 Report for the 20th Class Day Census)

Access Database



Reporting & Procedures Manual

Univ. Student Report (CBM001)

DATA PROCESSING RECORD LAYOUT

		Beginning Position	Length
Item #1	Record Code - Always '1'	1	1
Item #2	Institution Code - FICE - Numeric	2	6
Item #3	Student Identification Number	8	9
Item #4	Gender - 'M' or 'F'	17	1
Item #5	Classification - Alphanumeric	18	1
Item #6	Date of Birth - YYYYMMDD - Numeric	19	8
Item #7	Tuition Status - Alphanumeric	27	1
Item #8	Residence - Numeric	28	3
Item #9	Transfer in College - Numeric or blank	31	6
Item #10A	SCH Load - Leading Zeros, two decimals	37	4
Item #10B	Unused - Zero fill	41	4
Item #11	Doctoral Hours Funded - Leading Zeros, two decimals	45	4
Item #13	Semester - Numeric	49	1
Item #14	Year - YYYY - Numeric	50	4

CBM001 Access Shell

The screenshot shows the 'CBM001_Submission' table structure in Microsoft Access. The table has two columns: 'Field Name' and 'Data Type'. The fields are listed in a table below.

Field Name	Data Type
1	
FICE	Short Text
SSN	Short Text
SEX	Short Text
CLS	Short Text
DOB	Short Text
TUIT ST	Short Text
RESD	Short Text
TRF/FT	Short Text
SCH ON C	Short Text
SCH OFF C	Short Text
DR HRS F	Short Text

Below the table structure, there are tabs for 'General' and 'Lookup'. The 'Field Size' property is set to 6, and the 'Format' property is empty.

Process – Access File Cleanup # 1

(For this presentation I will focus on the CBM001 Report for the 20th Class Day Census)

Confluence - CBM Report Notes & Checklists

Dashboard / ... / CBM001

Fall 2020 CBM001/CBM004

Created by Farrell, Liza, last modified on Sep 28, 2020

Resources:

- CBM Reporting Manuals
- TAMUCC Course Inventory
- Program Inventory
- Instructions to update FE Entry Courses

CBM Data Management Log - File to enter CBM errors, review root causes and resolutions previously applied. All directs you to the right functional lead. (\\VAPA\THECB\CBM Reports\CBM Documentation)

Notes

FE students with an FE of 3 are Doctoral students with multiple funding codes and do not actually count as FE sts

Action Items

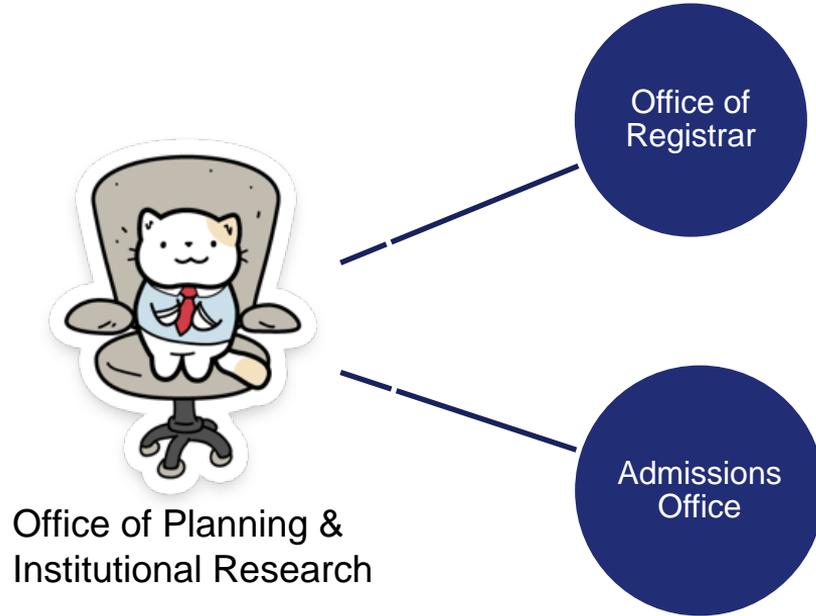
- Check FE courses list to update on SZASXRF: @Merchant, Amanda
- Update FE courses on SZASXRF in Banner @Farrell, Liza 20 Aug 2020
- Run SSR01.ST query for Cross-Listed courses. @Farrell, Liza 22 Sep 2020

Access Database – CBM001 Report

	TUIT ST	RESD	TRF/FT	SCH ON C	SCH OFF C	DR HRS F	SEM	YEAR	SCH II F	FE	RC	MAJR CIP
1	205	000001	1300	0000	0000	1	2020	0000				26010200
1	125		1200	0000	0000	1	2020	0000				26010200
1	178	003563	1600	0000	0000	1	2020	0000				42010100
1	014		1500	0000	0000	1	2020	0000				26010100
1	125		1200	0000	0000	1	2020	0000				52020100
1	125	000001	1100	0000	0000	1	2020	0000				51380100
1	125		1300	0000	0000	1	2020	0000				26010200
1	125		1300	0000	0000	1	2020	0000				42010100
1	125		1500	0000	0000	1	2020	0000				42010100
1	108		1200	0000	0000	1	2020	0000				51380100
1	205	000001	1300	0000	0000	1	2020	0000				51380100
1	066		1500	0000	0000	1	2020	0000				50040900

Process – Access File Cleanup #2

(For this presentation I will focus on the CBM001 Report for the 20th Class Day Census)

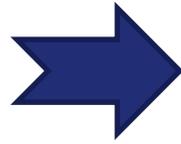


Process – Certification

(For this presentation I will focus on the CBM001 Report for the 20th Class Day Census)



Office of Planning &
Institutional Research



THECB CBM EDC
Portal Submission



Certification is
COMPLETE



Append

Addressing Challenge 3

Appending the data into historical tables in SQL Server

SQL- Historical CBM tables overview

- TAMUCCPIR:
 - Historical_cbm001
 - Historical_cbm002
 - Historical_cbm004
 - Historical_cbm008
 - Historical_cbm009
 - Historical_cbmb
 - Historical_cbms
-

Step 1: Exporting it from Access to Excel



Step 2: Manual Mapping

	SSN	PIDM	BIRTH_YEAR	BIRTH_MONTH	SCH	TERM
1	X1	4532	1995	9	9	20209
2	X2	3452	1993	2	12	20209
3	X3	5653	1997	5	6	20209
4	X4	2678	1994	10	9	20209

SQL historical_cbm001 Table

Data to append to SQL historical_cbm001

A	B	C	D
SSN	DATE_OF_BIRTH	SCH	PIDM
Y1	19960305	0900	4343
Y2	19930907	1200	2534
Y3	19980203	0600	1256
Y4	19911103	0800	5234

CBM1 Spring 2021

Difference between historical table and certified file

	SSN	PIDM	BIRTH_YEAR	BIRTH_MONTH	SCH	TERM
1	X1	4532	1995	9	9	20209
2	X2	3452	1993	2	12	20209
3	X3	5653	1997	5	6	20209
4	X4	2678	1994	10	9	20209

SQL historical_cbm001 Table

A	B	C	D
SSN	DATE_OF_BIRTH	SCH	PIDM
Y1	19960305	0900	4343
Y2	19930907	1200	2534
Y3	19980203	0600	1256
Y4	19911103	0800	5234

CBM1 Spring 2021

CBM1 Spring 2021 data after mapping manually

	A	B	C	D	E	F
1	SSN	PIDM	BIRTH_YEAR	BIRTH_MONTH	SCH	TERM
2	Y1	4343	1996	3	9	20211
3	Y2	2534	1993	9	12	20211
4	Y3	1256	1998	2	6	20211
5	Y4	5234	1991	11	8	20211

CBM1 Spring 2021 data ready to append

Mapping process automated using Python

```
import pandas as pd
import xlrd
import openpyxl
```



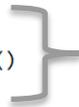
Importing required python Libraries

```
current_term = pd.read_excel('data.xlsx', dtype=str)
current_term = current_term.fillna('')
current_term.columns = ['ID', 'SSN', 'PIDM', '1', 'FICE', 'SEX', 'CLASS', 'DOB', 'TUIT', 'RES', '']
sql_append = pd.read_excel('20199.xlsx')
```



Reading Input files as data frames

```
sql_append = sql_append[0:0]
print(sql_append)
print(current_term.loc[:, ['DOB']])
current_term['BIRTHY'] = current_term['DOB'].str[:4]
current_term['BIRTHM'] = current_term['DOB'].str[4:6]
print(current_term['BIRTHY'])
print(current_term['BIRTHM'])
del(current_term['DOB'])
for i in sql_append.head():
    if i in current_term.head():
        sql_append[i]=current_term[i]
sql_append['TERM'] = '20199'
sql_append['tempID'] = ''
cols_lead_zeros = ['SCHON', 'SCHOFF', 'SCHDOC', 'SCH_NOT_FUND', 'SCH_DEV_NO_FUND', 'SCH_INTE']
for i in cols_lead_zeros:
    sql_append[i.lower()] = sql_append[i].apply(int)
    del(sql_append[i])
cols_arrangement = ['SSN', 'SEX', 'CLASS', 'BIRTHY', 'BIRTHM', 'TUIT', 'RES', 'TRANS', '']
for i in cols_arrangement:
    if i in cols_lead_zeros:
        cols_arrangement[cols_arrangement.index(i)] = i.lower()
sql_append = sql_append[cols_arrangement]
print(cols_arrangement)
print(sql_append)
sql_append.to_excel('output3.xlsx', index=False)
```



Arranging columns in the same order as in historical_cbm001



Output: Excel file ready to append

Input to Python program: Certified Spring 2021 Excel File

A	B	C	D
SSN	DATE_OF_BIRTH	SCH	PIDM
Y1	19960305	0900	4343
Y2	19930907	1200	2534
Y3	19980203	0600	1256
Y4	19911103	0800	5234



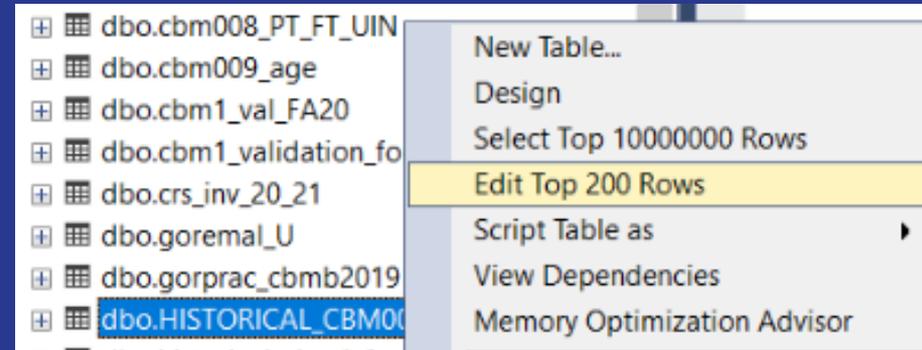
Python Program



Output from above python program: Excel File, with all changes done and ready to append

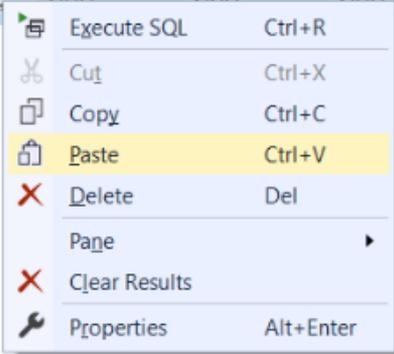
	A	B	C	D	E	F
1	SSN	PIDM	BIRTH_YEAR	BIRTH_MONTH	SCH	TERM
2	Y1	4343	1996	3	9	20211
3	Y2	2534	1993	9	12	20211
4	Y3	1256	1998	2	6	20211
5	Y4	5234	1991	11	8	20211

2: Open SSMS, right click on the table in which you want to append data to (in our example, the table is Historical_cbm001), left click on **Edit Top 200 Rows** option.



3: Go to the bottom and right click on the left most cell with a * in the last row and hit **Paste**

	SSN	PIDM	BIRTH_YEAR	BIRTH_MO...	SCH	TERM
	X1	4532	1995	9	9	20209
	X2	3452	1993	2	12	20209
	X3	5653	1997	5	6	20209
	X3	2678	1994	10	9	20209
				NULL	NULL	NULL



- Execute SQL Ctrl+R
- Cut Ctrl+X
- Copy Ctrl+C
- Paste Ctrl+V**
- Delete Del
- Page
- Clear Results
- Properties Alt+Enter

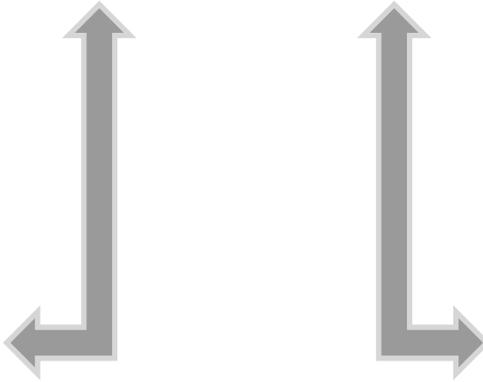
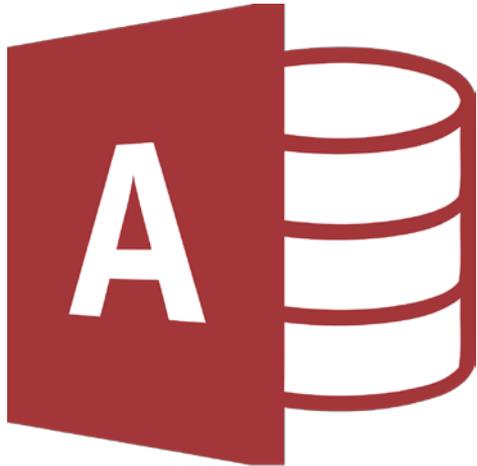
Accessing the newly appended **TERM** data

```
/****** Script for SelectTopNRows command from SSMS *****  
SELECT TOP (10000000) [SSN]  
    , [PIDM]  
    , [BIRTH_YEAR]  
    , [BIRTH_MONTH]  
    , [SCH]  
    , [TERM]  
FROM [data_cleaning].[dbo].[HISTORICAL_CBM001]  
where TERM = '20211'
```

119 %

Results Messages

	SSN	PIDM	BIRTH_YEAR	BIRTH_MONTH	SCH	TERM
1	Y1	4343	1996	3	9	20211
2	Y2	2534	1993	9	12	20211
3	Y3	1256	1998	2	6	20211
4	Y4	5234	1991	11	8	20211



Impact

Fewer errors

Faster turn around time

Growth in campus trust

Increase in data requests

Foundation for analytics

Ability to invest in research



Question

How can we best build synergies and open communication lines, not only in our own team, but within TAIR overall?

Contacts

Amanda Merchant: amanda.merchant@tamucc.edu

Liza Farrell: liza.farrell@tamucc.edu

AbdulHai Mohammed: abdulhai.mohammed@tamucc.edu