



explorance
world 2024



Everything is Changing: Working Through the Data Weeds

Eowyn Ellison – University of Maryland

Agenda

- Why is everything changing?
- Overview of the planning process
- Structure, Design and Testing
- Set-up For Full Testing
- Updating Scripts
- Where Do We Go Now?
- Questions

Why is everything changing?

- Our data structure is entirely built on the internal structure of our HR system
- Elevate (Workday) - completely replace and restructure HR and Finance now, Student Information in the future

Planning Process

- We had to know to handle the new structures
- We had to test usability of the new structures in Blue
- We had to do a full test in Blue
- We needed to delve into how the tables used for Blue are built, maintained, and their interdependencies
- We need to know how things will be maintained between HR GoLive and Student GoLive

Structure, Design and Testing

- We used the sandbox warehouse set up of data structure to understand how the new structures work
- We used the draft mapping from the sandbox as a good approximation for what we would deal with with the various structures
- We had to decide which of the 3 structures we would use (decision - academic hierarchy)

Structure, Design and Testing

- The structure in current state is set up so that there is a Unit -> Department -> College hierarchy where the code for the unit includes codes for department, college (and higher levels)
- A course prefix is assigned to a unit and it will always have a department that owns the unit (and thus the prefix) and a college that owns the department

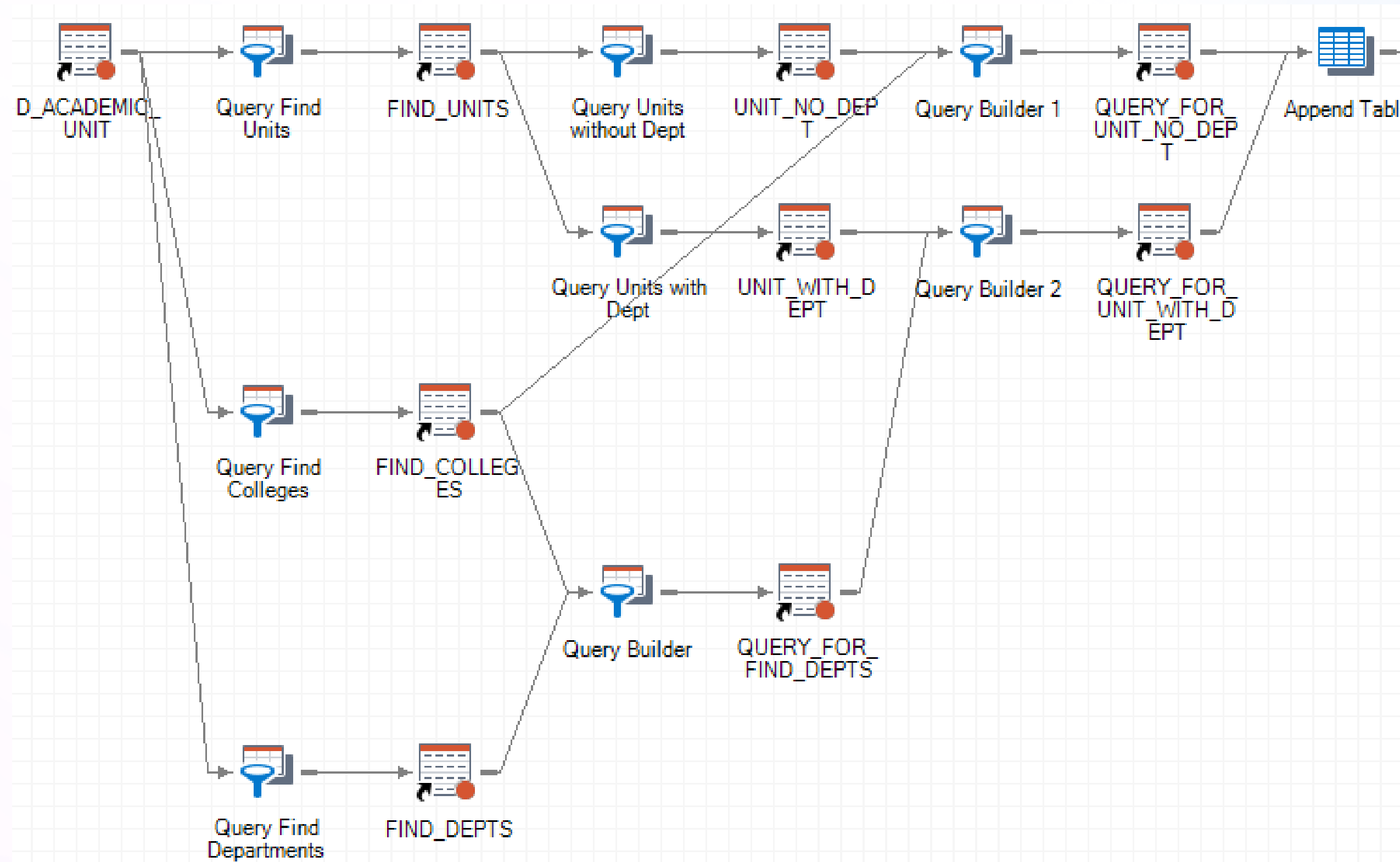
Structure, Design and Testing

- To control for variability, we pulled a ‘master’ set of the new structure to use in testing.
- We did a test text-map of the string values for colleges and departments from new onto old to see how easy the mapping would be.
 - some of the hierarchy pieces had holes
 - some of the names were slightly different (& vs. and vs. And)
 - exercise will be used in production to build the course prefix to entity map

Structure, Design and Testing

- The structure in future state is Child-Parent and what the designation of each (sub-unit, unit, department, college, university, etc.) is independent.
 - we will have prefixes owned by any of the entities, not just units
 - structure will not work in the Blue hierarchy rules without using multiple hierarchies
 - we can find a list of departments and colleges using coding in the structure for what type of entity and find the holes but it is messy

Structure, Design and Testing



Structure, Design and Testing

- We are creating calculated data for the department gaps

`IFC(t1.Dept_Ref_ID IS MISSING,(CATS('DEPT_',t1.Col_Ref_ID)),
t1.Dept_ref_id)`

i.e. if there isn't a department in the string, make the department code be Dept_[college code]

- We are deciding how to deal with certain situations and starting to code them.

Structure, Design and Testing

- With initial replacement structure we could start the first testing
 - upload new structure and create a project with triggers
 - do a quick test of data collection and reports

Set-up For Full Testing

- We are going to have to switch structures during data collection
- We wanted to check a variety of different scenarios
 - all merge possibilities (cross-list, lecture, etc.)
 - all trigger possibilities
 - changing triggers in a live project
 - changing data structure in a live project
- This would have to be a 2 round test

Set-up For Full Testing

- Find a set of courses that would fulfill what was needed
- Create a map of old->new structure

Blue Triggers			
Level	Meaning	Value	Changed Value
Col	Policy	PLCY	
Dept	PSYC	12825	AU00209
Dept	ENGL	12719	AU00170
Dept	Materials Sci	13219	AU00323

Set-up For Full Testing

- I created a spreadsheet with 1 row/course section and columns for the testers in Round 1 and Round 2
 - each course had people in both rounds (with one paired exception)
 - each tester had courses in both rounds, trying to balance the numbers
 - each tester was in each course a maximum of 1 time
- Each tester was given a guideline for each course (e.g. like, dislike, do a pattern, be middling) to cover all possibilities

Set-up For Full Testing

- Use that master to create an updated Student-Course file with all the courses for testing

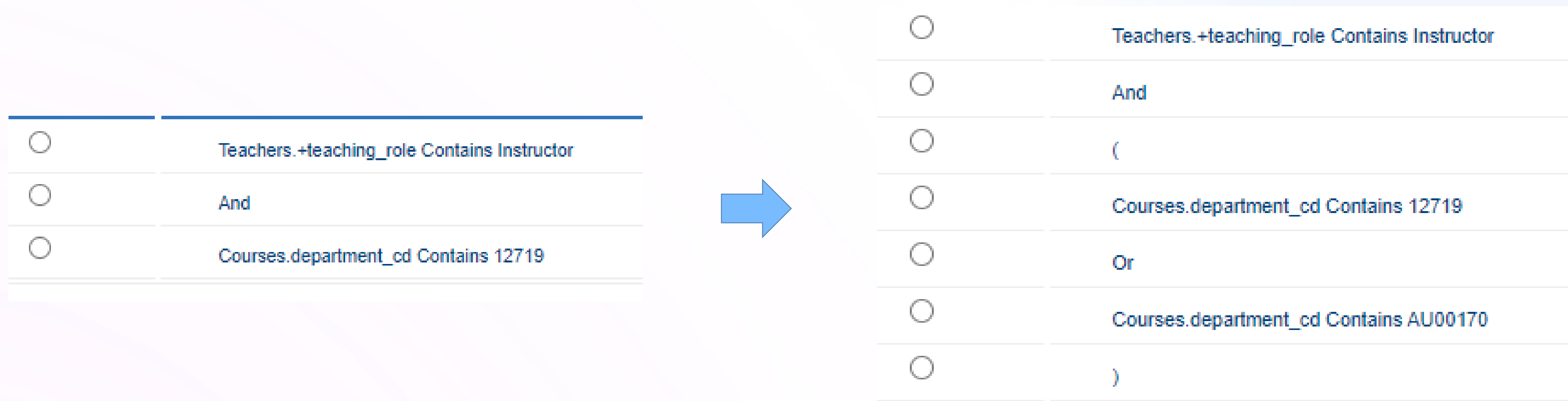
courseid	studentid	termid	mod_dt
202308-BSCI338N-0101	999888777	202308	#####
202308-BSCI338N-0101	888777666	202308	#####
202308-BSCI338N-0101	777666555	202308	#####
202308-BSCI338N-0101	666555444	202308	#####
202308-BSCI338N-0101	555444333	202308	#####
202308-BSCI338N-0101	444333222	202308	#####
202308-BSCI338N-0101	333222111	202308	#####
202308-BSCI338N-0101	999777555	202308	#####
202308-BSCI338N-0101	888666444	202308	#####
202308-BSCI338N-0101	777555333	202308	#####
202308-CMLT398L-0101	999888777	202308	#####
202308-CMLT398L-0101	888777666	202308	#####
202308-CMLT398L-0101	777666555	202308	#####

Set-up For Full Testing

- We needed a test as close to what will happen as possible
 - I launched the project with the Student-Course file using a copy of a previous project
 - We had a small group of people do some submissions
 - Then update the triggers and have the rest of Round 1 submissions done
 - By tracking the number of submissions, I was able to figure out where a mistake was made, fix it and continue

Set-up For Full Testing

- Since we had people submitting on the old structure while the new triggers were in place we had to include both structures in the triggers.



Set-up For Full Testing

- After all of Round 1 was submitted, we needed to pause in order to do the updates.
 - upload new Hierarchy file
 - upload new Report Viewer file
 - upload new Course file
- Then we did a manual sync of the affected fields

<input type="checkbox"/>	Field Name	Automatic Sync	Added Date	Last Updated
<input type="checkbox"/>	course	No	2/12/2024 4:20:28 PM	
<input type="checkbox"/>	course_prefix	No	2/12/2024 4:20:28 PM	
<input checked="" type="checkbox"/>	department	No	2/12/2024 4:20:28 PM	2/26/2024 12:14:30 PM
<input checked="" type="checkbox"/>	department_cd	No	2/12/2024 4:20:28 PM	2/26/2024 12:14:30 PM

Set-up For Full Testing

- We had a subset of users go in and check triggers.
- Then we had the rest of the testers complete Round 2.
- I updated the triggers in a report and ran it.
- I checked the generated reports (with only the new triggers) for display of data from both rounds.

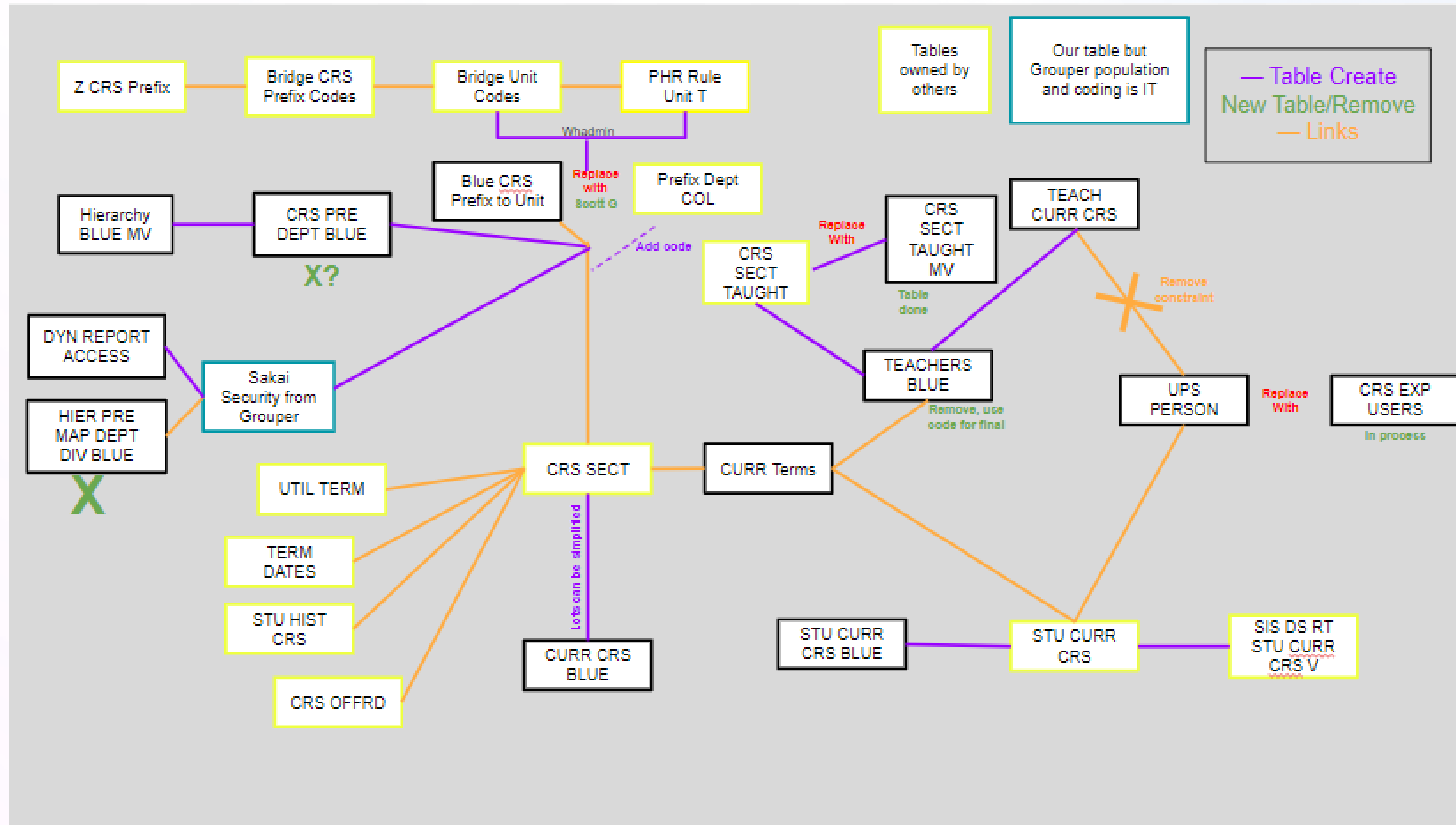
Updating Scripts

- We have copies of the SQL for several tables and asked for the others.
- I then went through each script to see what the source tables are so we could see the intersections beyond the original data flow that was designed in 2016.

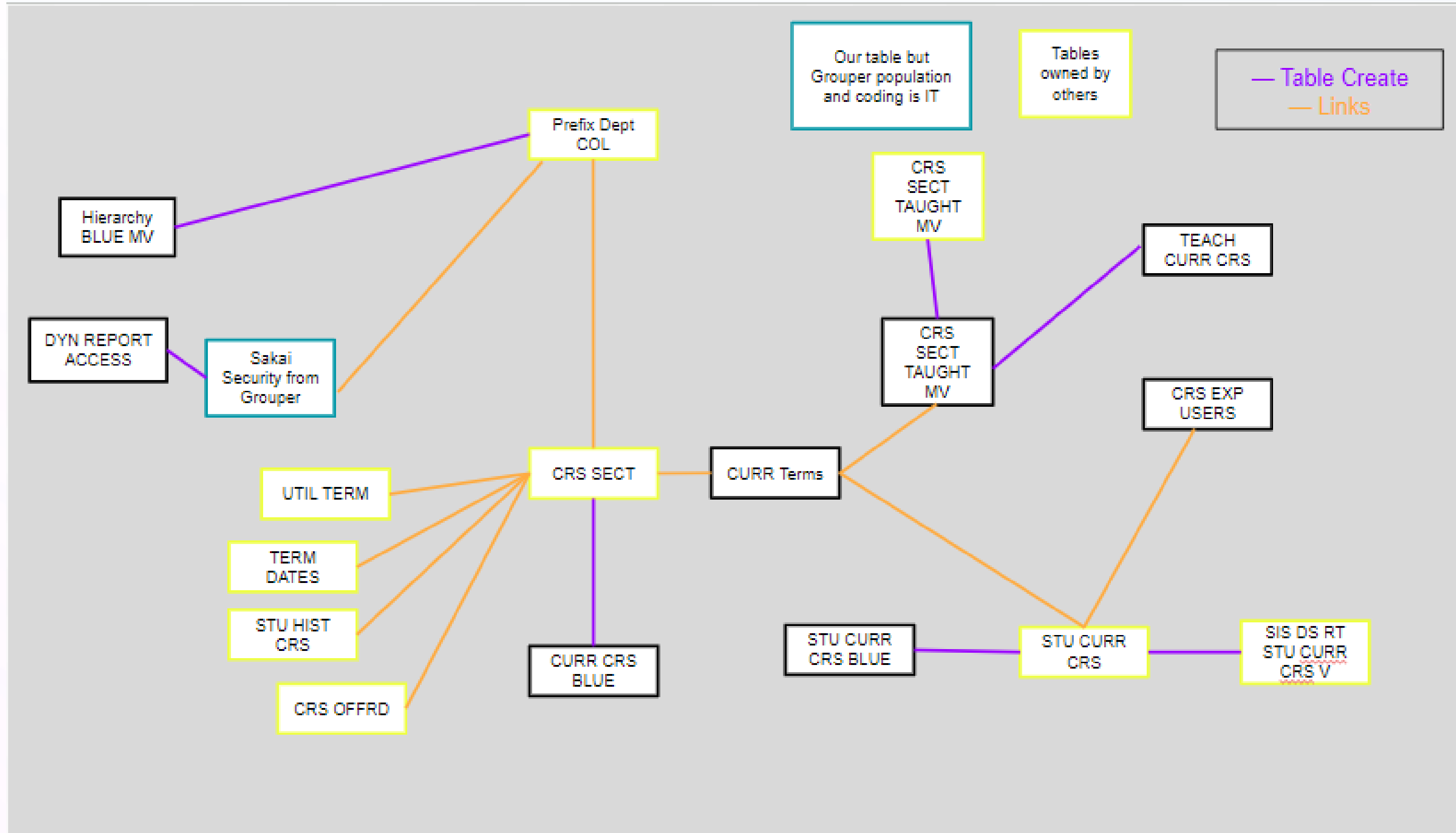
Updating Scripts

- We looked at the code that generates the tables we push to Blue looking for ways to make them more efficient
 - The User table depended on tables that are being deprecated, that has been updated.
 - The Course-Teacher table depended on a view that was becoming too large, we now have a Materialized View for that.
 - The Course table depends on 16 tables, we are dropping it to 6 (eventually to 5).
 - Our Dynamic Report Viewers depends on 6 tables, dropping to 2.
 - We are documenting all the source tables in each script because we will do this again when Student goes live.

Updating Scripts



Updating Scripts



Updating Scripts

- The scripts have all been drafted and are with the DBAs, as each table is built in QA, I will check them.
- Once all of the updated scripts are working and stable, we will build a few projects in the Blue Sandbox.
- Once that is done, we will have our IT partners update the DataSync tool to read from the QA warehouse and push to the Blue Sandbox (not impacting Prod).

Where Do We Go Now?

- We are working with a group of stakeholders to create bridges between current state and future state.
 - Some of the bridges will be for a one time map from one to the other
 - Some of the bridges (including prefix to entity) will be more of a dimension table
- We are also working to determine how the bridge/dimension table will be maintained.
 - How will the curriculum people add prefixes.
 - How will they update ownership.
 - How will they flag deprecated prefixes.

Where Do We Go Now?

- Once the final version of the academic structure is created (before HR GoLive), build the initial map of prefixes to entities and get it approved by the data steward.
 - This will be used for the initial bridge/dimension table.
- After that is done, the production version of the new warehouse can be built with that map and the final academic structure.
- Our scripts will be updated in QA to read from the production build for testing.

Where Do We Go Now?

- We will update the triggers in the active project to the new structures and triple check.
- We will update triggers in all versions of the old reports (changes in triggers, changes in questions).
- We will work with the DS people to pause while scripts are flipped and checked in Prod, then do a manual sync and test (very fast) on a copy of the project.
- At that point we can begin updating all of the old projects and rebuilding and running the reports oldest to newest.

Questions?

Eowyn Ellison
University of Maryland
eowyn@umd.edu

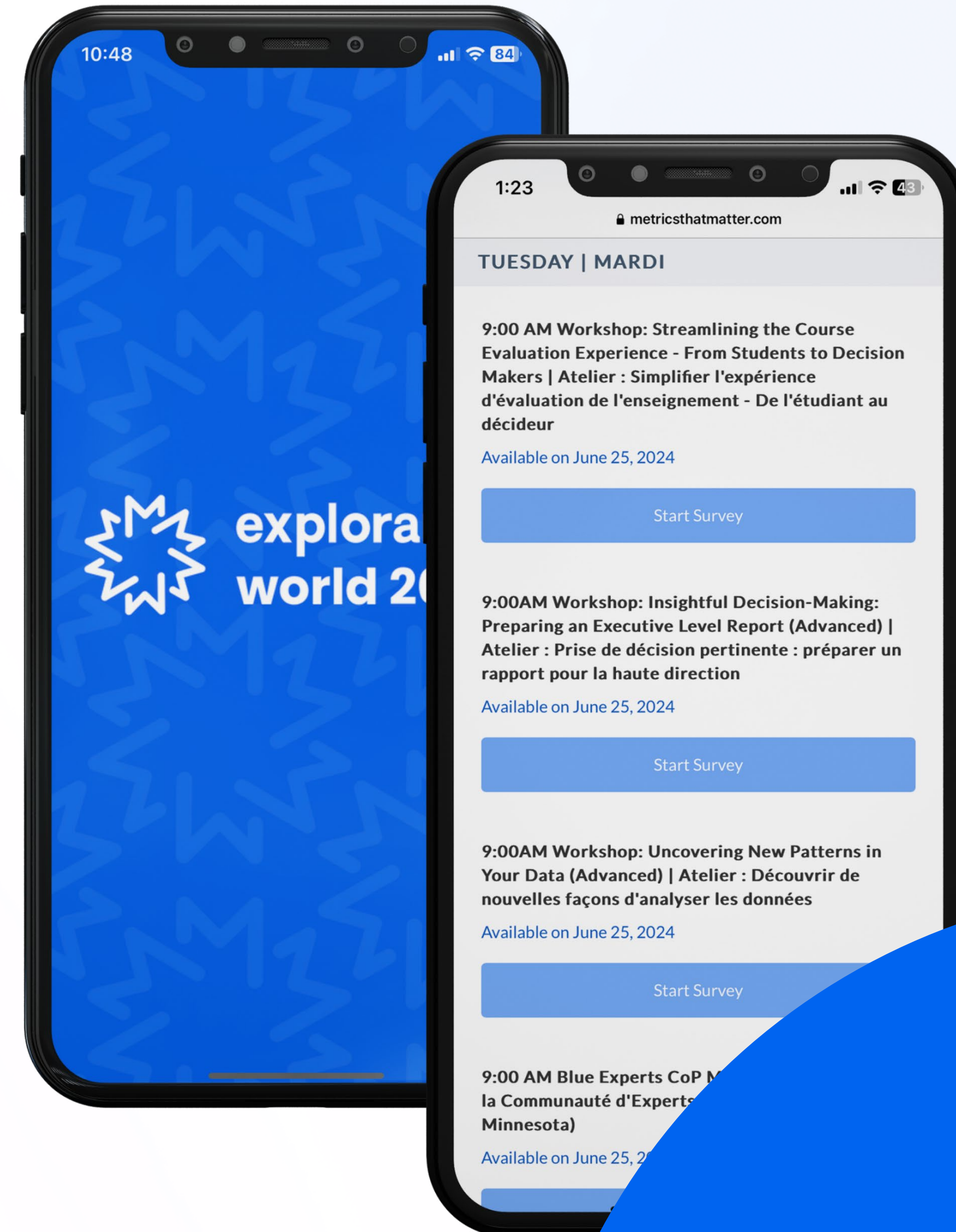


explorance
world 2024

Provide Your Feedback!



CONFERENCE SURVEYS



 **Everything is Changing: Working Through the Data Weeds**

 **Everything is Changing: Working Through the Data Weeds**

 **Everything is Changing: Working Through the Data Weeds**

 **Everything is Changing: Working Through the Data Weeds**