education
analytics

Interoperable Data for Research: Texas Education Exchange

Molly Stewart[1], Amy Musgrave[2], and Andrew Rice[1]

1: Education Analytics
2: Texas Education Agency

Amy Musgrave
TEA

Dr. Molly Stewart
Education Analytics

Dr. Andrew Rice
Education Analytics

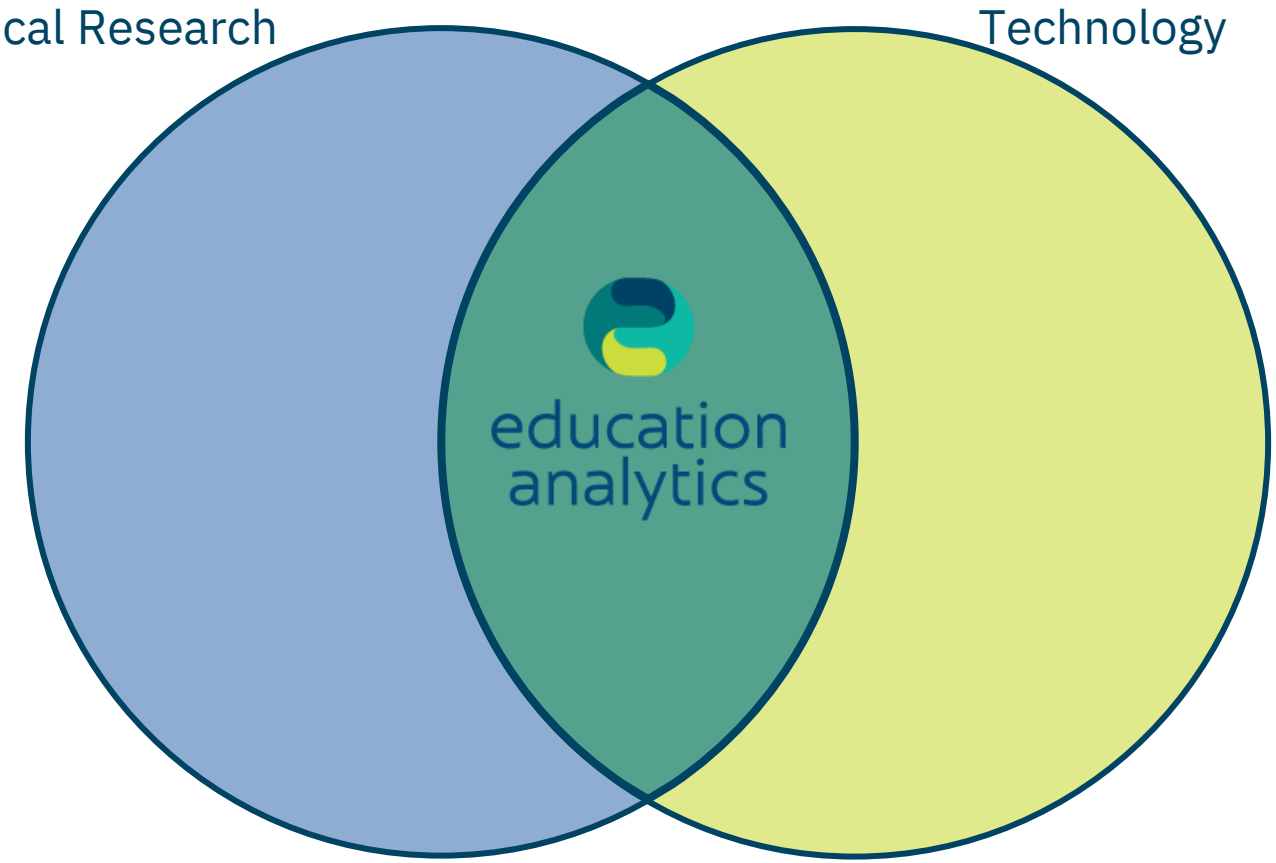

How can we make research cheaper, faster and more actionable using modern data technology?

Three Ideas

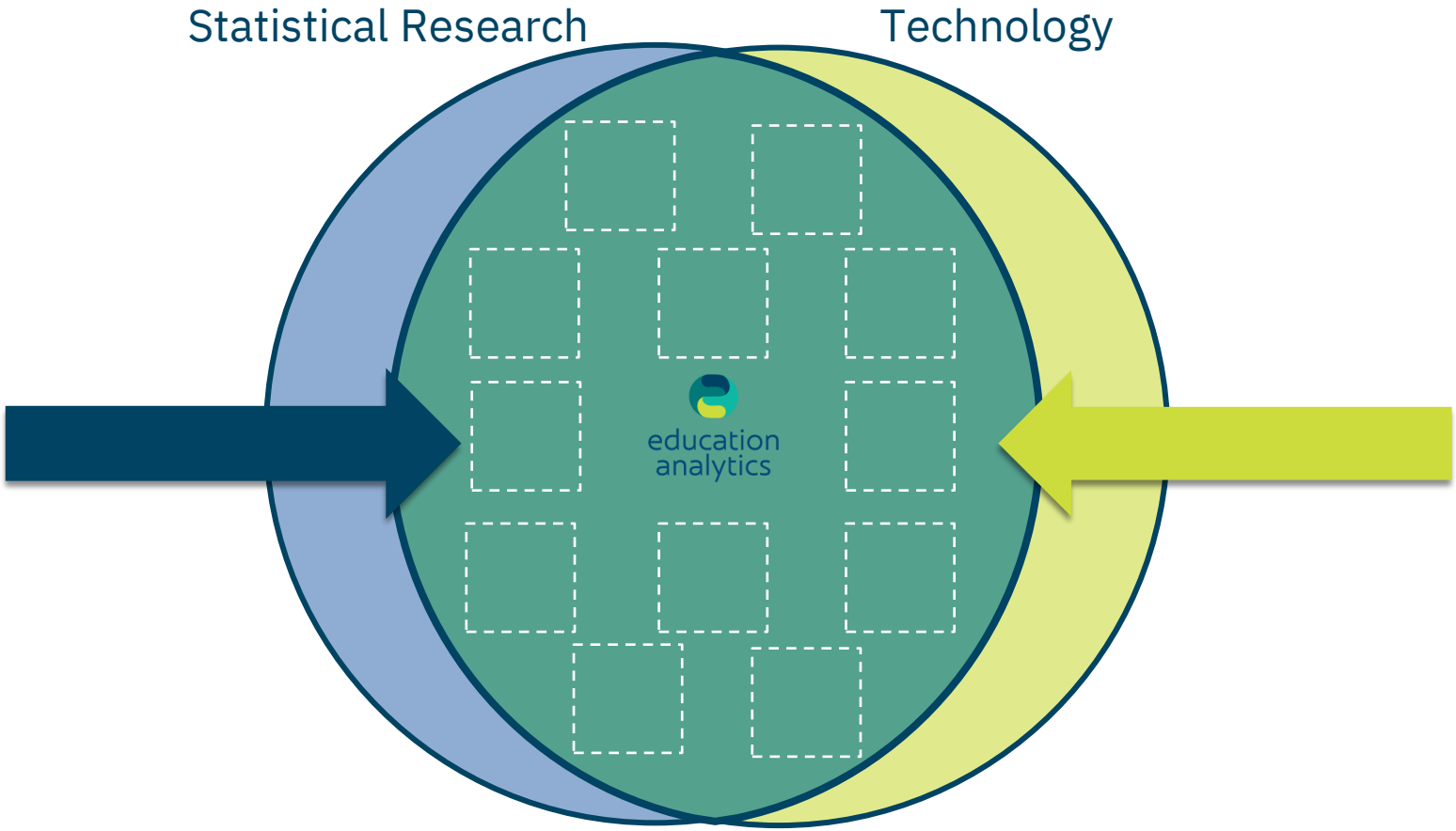
Takeaways

- Data interoperability technology is a step function increase in data availability for all research
- New research questions can be asked and answered at scale and replicated in every location
- The TEA with the Texas Education Exchange is using this technology in Texas as we speak

Who is EA?



Who is EA?



Who is EA

- 501(c)(3) national nonprofit based in Madison WI
  - Founded by colleagues at the Wisconsin Center for Education Research (WCER) at the University of Wisconsin
- 120ish researchers, data analysts, software developers, data engineers, technologists, and support staff
- National scope working with hundreds of district and state departments of education
- **Primary data technology partner of the Texas Education Exchange**

What is the Texas Education Exchange?

- Amy Musgrave, Director of Digital Strategy
- In 2023, the Texas Education Agency began the process of implementing an interoperable system for data collection (statewide data warehouse)
  - Region 4 created the Texas Education Exchange to offer local data warehouses using the same data structures as the TEA statewide warehouse
  - Districts can utilize this data warehouse, BI dashboards, and a teacher-facing application, and other tools by opting in
  - Districts can also give permission for vendor or researcher data access to data, easing the burden for dataset creation

LEAs Control Their Data and Utilize It for Local Use

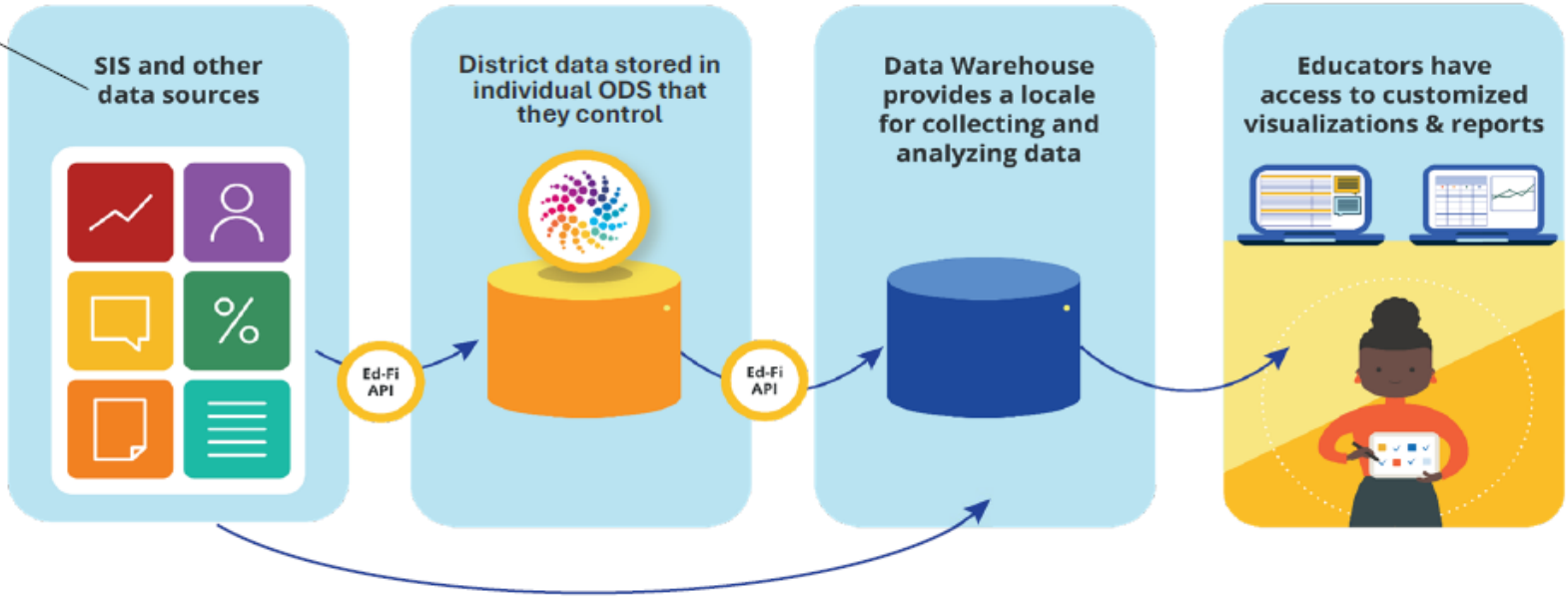
Anyone can build an application with an Ed-Fi API that LEAs connect to their Exchange environment

Examples of Data:

Student data
- Identification
- Demographic
- Enrollment
- Attendance
- Special Program
- Alternative &
- Supplemental Services
- Discipline
- Etc.

Staff Data
- Staff Assignment
- Finance-payroll
- Courses offerings
- Etc.

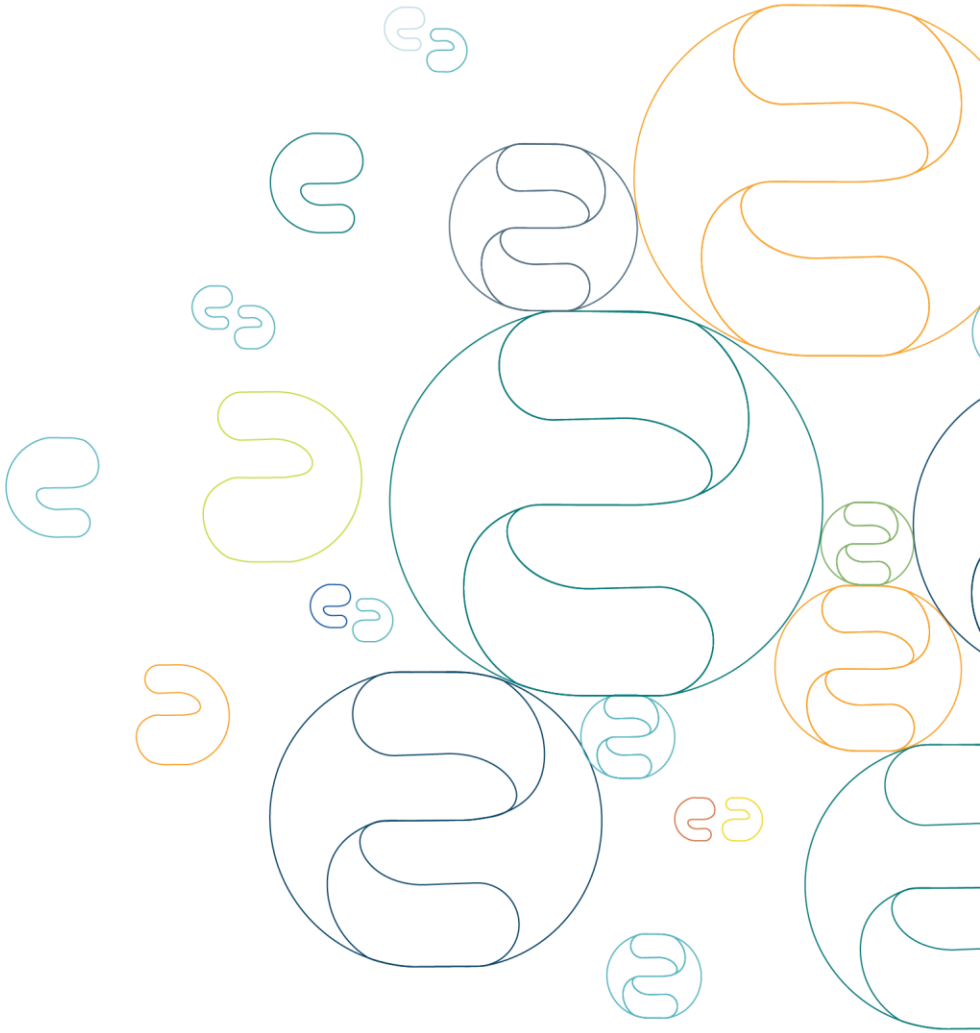
Assessment Data
- Benchmark
- Interim
- Local



SIS = Student Information System (like ASCENDER)
ODS = Operational Data Store

Participating Districts

- Over 150 districts in Texas currently have data in this warehouse – many are smaller, rural districts
- By SY25-26, over 900 of Texas's 1250 LEAs will be included



What is Interoperability?

- The ability to flow data between two or more systems using a common data standard (format) using an API
- Data formats from source systems are transformed into the common data standard in the receiving system
- For research purposes, this enables
  - Repeatable data cleaning and querying instead of receiving, cleaning, and merging a number of unique files from research partners
  - Transparency into the application of business rules (e.g., which data are included or excluded in a dataset)
- Data standards for interoperability can vary in design

What is Ed-Fi?

- The interoperability technology and data standard that TEA and the Exchange use
- Transactional data from as many systems as possible
  - SIS (attendance behavior, grades)
  - Assessment
  - HR/staffing
- Think: if it is being recorded in a tech system it will eventually be integrated into the Exchange

How in the world does it work?

- Think like a technologist:
- We can build a “research application” that allows for data access for research
- This can control permissions and also manage data release agreements
- **DOES NOT CHANGE THE LEGAL PROCESS**
  - **Districts still control every single data point and access level**

LEAs Control Their Data and Utilize It for Local Use

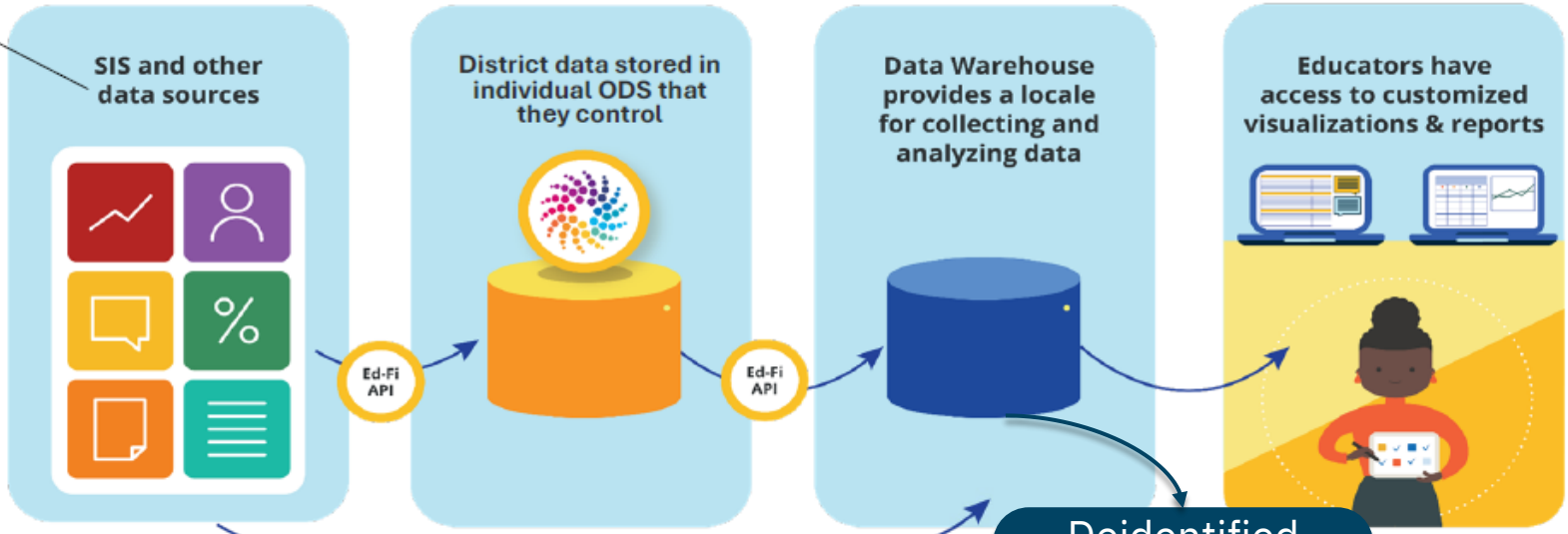
Anyone can build an application with an Ed-Fi API that LEAs connect to their Exchange environment

Examples of Data:

Student data
- Identification
- Demographic
- Enrollment
- Attendance
- Special Program
- Alternative &
- Supplemental Services
- Discipline
- Etc.

Staff Data
- Staff Assignment
- Finance-payroll
- Courses offerings
- Etc.

Assessment Data
- Benchmark
- Interim
- Local



Deidentified layer of warehouse (in development)

Research use (create data sets via simple SQL tables/queries)

SIS = Student Information System (like ASCENDER)
ODS = Operational Data Store

So what???

- This sounds like an IT project right?
- Wrong!
- Imagine a common dataset structure for all districts in Texas (one for each district) that is updated in real time and has all of the data
- What would you do?

Ed-Fi for Research/Evaluation

- Granularity
  - Attendance: one record per day per student (present/absent)
    - Reason detail for excused absences
    - Sometimes we get period absences, sometimes not
  - Discipline: Date, time of day, behavior category, staff person who reported it, action taken, and length of action (if >1 day)
    - Sometimes we get qualitative descriptions of the incident
  - Grade updates can be sent in real-time but more commonly we get them at the end of every grading period (still more frequent than end-of-term/transcript only)
  - Course-taking patterns, course pass/fail rates, class size
    - Linkages between student--course section--teacher
  - Time spent on assessments; percent correct; sub-section scores (depending on assessment)

Ed-Fi for Research/Evaluation

- Time-based analyses
  - Linking attendance, discipline, and achievement patterns over time
    - E.g.: Do absences follow disciplinary events? How does achievement change with absences and/or disciplinary events?

- Timeliness:
  - Districts often get important data from state agencies far too late to do anything about it (e.g., discipline disproportionality and reporting patterns)
  - Data updating in real-time or regularly during the school year offers opportunities to inform educators about what they can be doing differently before the year is over

API and Operational Data Store

What is JSON and what does it look like?

(Helpful to recognize, even if you don't need to do anything with it.)


```
[] V

{

"attendanceEventCategoryDescr
iptor": "uri://ed-
fi.org/AttendanceEventCategor
yDescriptor#Excused Absence",
  "eventDate": "2023-09-07",
  "schoolReference": {
    "schoolId": 9999105
  },
  "sessionReference": {
    "schoolId": 9999105,
    "schoolYear": 2024,
    "sessionName": "9999105-
S1"
  },
  "studentReference": {
    "studentUniqueId":
"104564"
  }
}
```


studentUniqueId	schoolId	school year	event Date	AttendanceEventCategoryDescriptorId
104564	9999105	2024	2023-09-07	431

DescriptorId	codeValue
431	Excused absence
432	Unexcused absence

Then you'd have to join these and other tables together, etc. – not very user friendly for research.

API and Operational Data Store

```
WITH SchoolYearDate AS (
        SELECT '2020-06-05 00:00:00.000' AS Date -- right here
), SchoolYear AS (
SELECT TOP 1 SchoolYear
        FROM edfi.SchoolYearType
        WHERE CurrentSchoolYear = '1'
), NumberDaysBySchool AS (
        SELECT
                SchoolId
                ,NumDays
                ,Date AS DayStart
                ,DATEADD(day,-1, LEAD(Date) OVER (PARTITION BY SchoolId ORDER BY Date ASC)) AS DayEnd
        FROM (
                SELECT DISTINCT
                        SchoolId
                        ,CCE.Date
                        ,DENSE_RANK() OVER (PARTITION BY SchoolId ORDER BY CCE.Date DESC) AS NumDays
                FROM edfi.[CalendarDateCalendarEvent] CCE
                CROSS JOIN SchoolYearDate AS SchoolYearDate
                JOIN edfi.CalendarEventDescriptor AS CED ON (CCE.CalendarEventDescriptorId = CED.CalendarEventDescriptorId)
                JOIN edfi.Descriptor AS CD ON (CD.DescriptorId = CED.CalendarEventDescriptorId)
                WHERE CCE.Date <= SchoolYearDate.Date
                        AND CD.CodeValue IN (
                                'Make-up day'
                                ,'Student late arrival/early dismissal'
                                ,'COVID-19 E-Learning'
                                ,'Instructional day'
                                , 'Make-up day'
                                , 'Delay/Early Dismissal - Other'
                                , 'Delay/Early Dismissal - Weather'
                                , 'E-Learning'
                                , 'E-Learning in Lieu of Cancellation'
                                , 'E-Learning PD'
                                , 'Instructional Half Day'
                                , 'Student Calendar'
                        )
```

Querying Ed-Fi Data in the Warehouse

Compare the previous script (page 1 of 4) to the one line of code that you need to type into the data warehouse to get year-to-date attendance data:

ANALYTICS.PROD_WH Settings

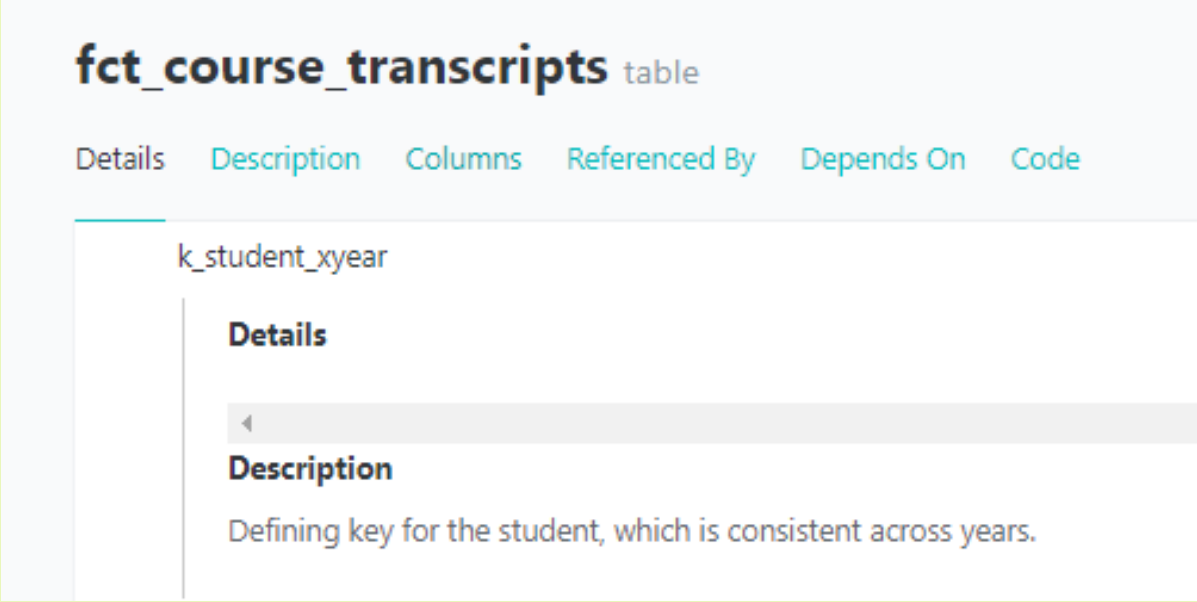
```
1  select * from fct_student_daily_attendance
```

Results Chart

	CALENDAR_DATE	ATTENDANCE_EVENT_CATEGORY	IS_ABSENT	IS_PRESENT	IS_ENROLLED	CUMULATIVE_ATTENDANCE_RATE
61	2024-08-06	In Attendance	0	1	1	50.00
62	2024-08-07	In Attendance	0	1	1	66.67
63	2024-08-05	In Attendance	0	1	1	100.00

Querying Ed-Fi Data in the Warehouse

- The data warehouse includes multiple years allows you to join data across years using the **k_student_xyear** variable (from dbt docs with permissions):



Data warehouses…
APIs…
Come on this still seems like an IT project!

Why leverage this ecosystem for research?

Granularity of data allows for analyses down to student, event (attendance/discipline), course-taking patterns, class size by teacher, subject, and student, assessment variables (e.g., time spent on assessment by subject), and others.

Time-based analyses

- Linking attendance, discipline, and achievement patterns over time
- E.g.: Do absences follow disciplinary events? How does achievement change with absences and/or disciplinary events?

Timeliness

- Districts often get important data from state agencies far too late to do anything about it (e.g., discipline disproportionality and reporting patterns)
- Data updating in real-time or regularly during the school year offers opportunities to inform educators about what they can be doing differently before the year is over

Example 2: Predicting Postsecondary Opportunities

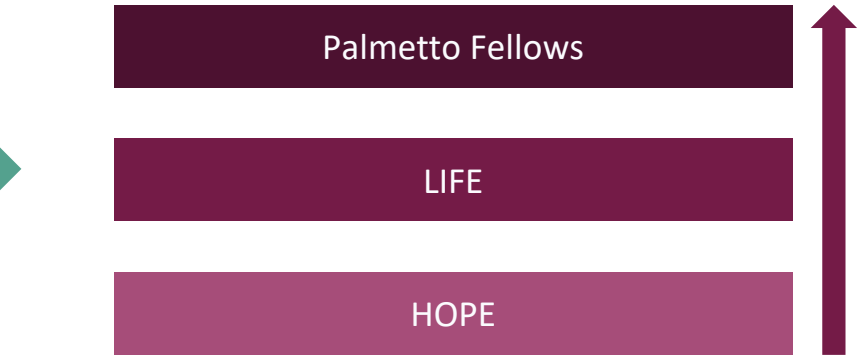
Goal: Predict the likelihood to be ready for postsecondary opportunities **early in a student's high school experience**, rather than waiting until the fall of a student's senior year to give them information about potential matches

Model Predicting Scholarship Eligibility

Ordered probit model predicting the likelihood of meeting the minimum requirements of three scholarships, ranked by selectivity level, using data from 2023-24 South Carolina graduating cohort

9th Grade Inputs:

- Attendance
- Course data
- End-of-course exams



Example 1: Peer effects of absences

Does peer absence on days absent matter?

- Compare absences for a student on a given day with other students in that course section on the same day

Results: No.

Ten percentage point increase in absence rate (reasons proportionate to statewide frequency) associated with reduction of:

0.51 letter grades when peer absence rate on days of own absence is 7 percent (1st quartile)

0.43 letter grades when peer absence rate on days of own absence is 17 percent (3rd quartile)

Effect mitigated to almost zero when student-level absence averages across courses are added to model. While effect is minimized, we can say it is measured very precisely with this dataset.

This study from start to finish took approximately 80 hours of work, including data cleaning. From our researcher:

"It would have been much more difficult, maybe even impossible, to do this without the data infrastructure. . . . Some of the data (like daily attendance by student) would have been just massive and unwieldy to have to import into statistical software and manipulate."

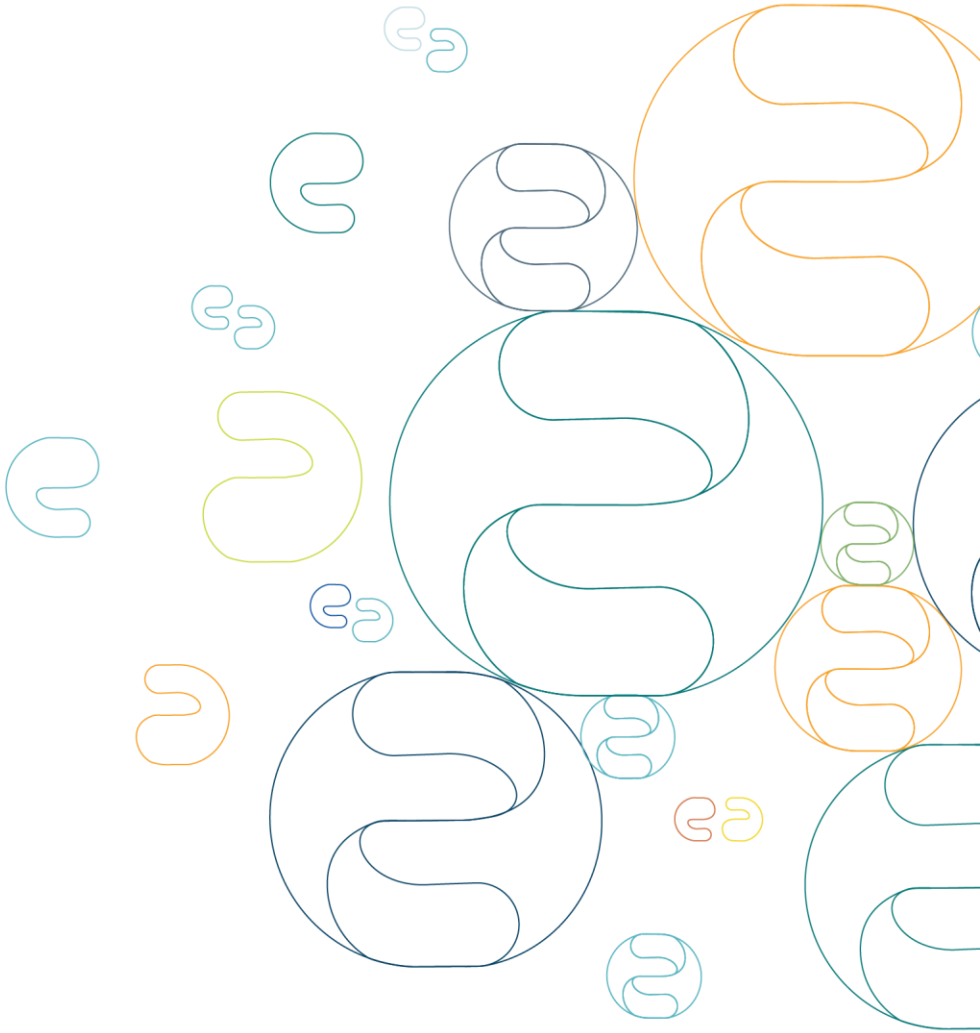
So what? We have all seen early warning and attendance research.

Here is what

- These projects were done at state scale with transactional level data in two weeks
- Done in SC but easily can be replicated in TN with “copy paste” of stats code
- Also – can be done on the Exchange with “copy paste” (if district want it)
- What about your quant research? What if you could do it once in your partner districts and replicate in every other district that wants it?

Discussion

- What can researchers do with this more granular, timely data?
  - EA blog about research questions that are enabled by Ed-Fi
- Permissions and data sharing agreements
- Districts currently participating in the Exchange (next slide)



Participating Districts

- Over 150 districts in Texas currently have data in this warehouse – many are smaller, rural districts
- By SY25-26, over 900 of Texas's 1250 LEAs will be included