Todays program:
Signe Løntoft (host and moderator); Altinget

Opening remarks:
Jonas Schytz Juul; Statistics Denmark and Fane Naja Groes; DRDS and Copenhagen Business School

Introducing 'Matched Educational Data' for general and vocational upper secondary education: Christian Vittrup, Asger Bromose Langgaard, and Alexander Erik Friisnæs; Statistics Denmark

Absence and Program Completion; Example of Data at Work:
Fane Naja Groes; Copenhagen Business School

Potentials of Matched Educational Data:
Andreas Bjerre Nielsen; University of Copenhagen

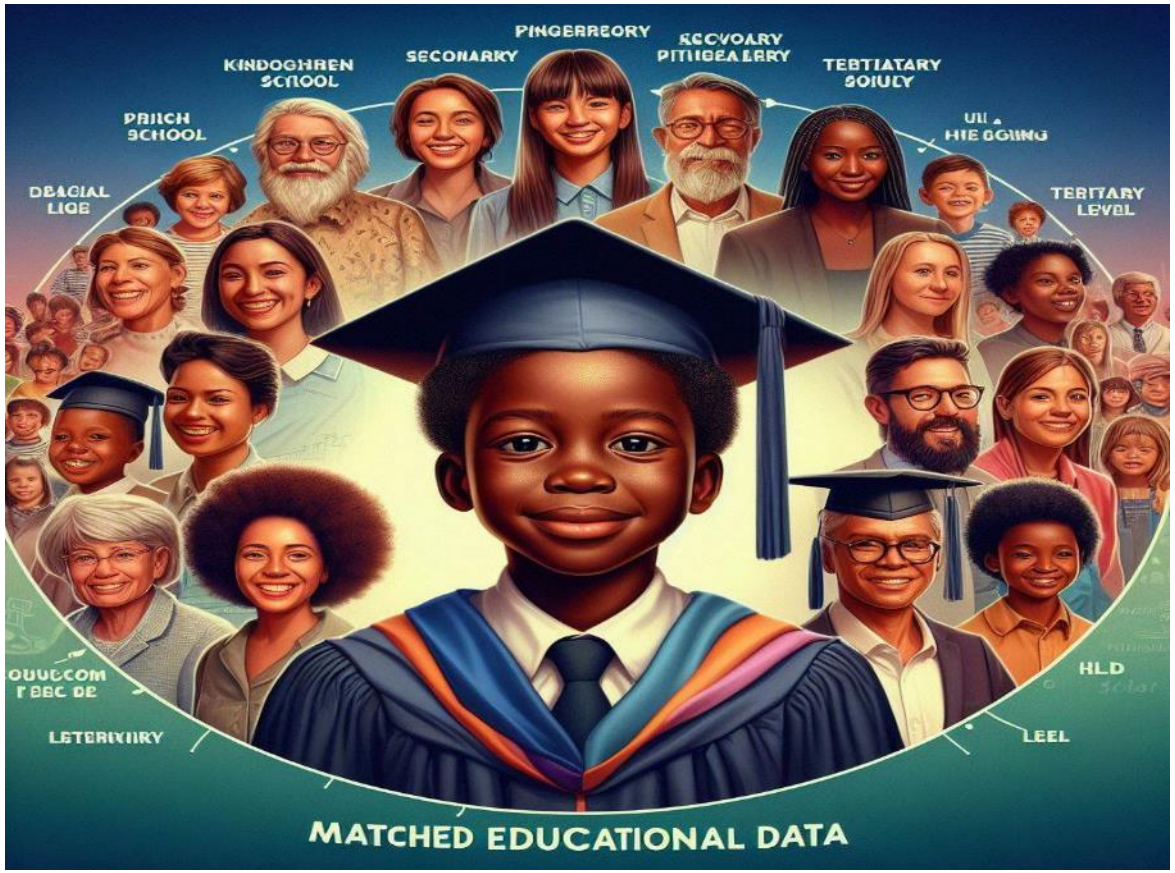
Data-Driven Change in the Educational Sector, a Panel Debate (in Danish) featuring:
Jesper Nielsen, Niels Ploug, Kristian Thor Jacobsen, and Jørgen Elmeskov

STATISTICS DENMARK

Matched Educational Data

Asger Bromose Langgaard,
Alexander Erik Friisnæs and
Christian Vittrup

Mail address for questions:
LER_Ungdomsuddannelser@dst.dk



STATISTICS DENMARK

I would like a photographic-style image depicting the developmental journey from kindergarten through primary school, secondary, tertiary education, up to the Ph.D. level. The image should include representations of individuals at various ages, symbolizing this educational journey, set against an educational-themed background. Additionally, please include a title that reads “Matched Educational Data”.

Intelligent Automation and Minimum Viable Product

STATISTICS DENMARK

Documentation

Scope

Automation

Integration

Frequency

Quality

DANMARKS STATISTIK

DST Quality metrics in register-based statistics

General Strategic framework for managing and mitigating errors in the data acquisition processes for registry based statistics

DANMARKS STATISTIK

- Follow UN's Generic Statistical Business Process Model GSBPM
- Preemptive error mitigation prior to data delivery
  - Enforce requirements on input formats
  - Macro-level frequency fluctuations
  - Flexible and close communication with data providers
- Extensive error handling process at Denmark Statistics
  - Macro-level frequency fluctuations across multiple levels
    - National/regional/municipal
      - Educational sector
        - Institutional
  - Micro-level error handling
    - Validation against reference registers and other data sources
    - Input errors
    - Inter-variable relationships
    - Rules and regulations
    - Frequent communication with data providers

Challenges of applying the general framework to the matched educational data

DANMARKS STATISTIK

- High frequencies data by few providers including all levels
  - No communication and troubleshooting on an institutional level
  - Not feasible to correct data retrospective
- More data on a micro level
  - Sheer volume of data is substantially larger
  - Complexity of inter-variable relations increases exponentially
  - Lack of pre validated reference data

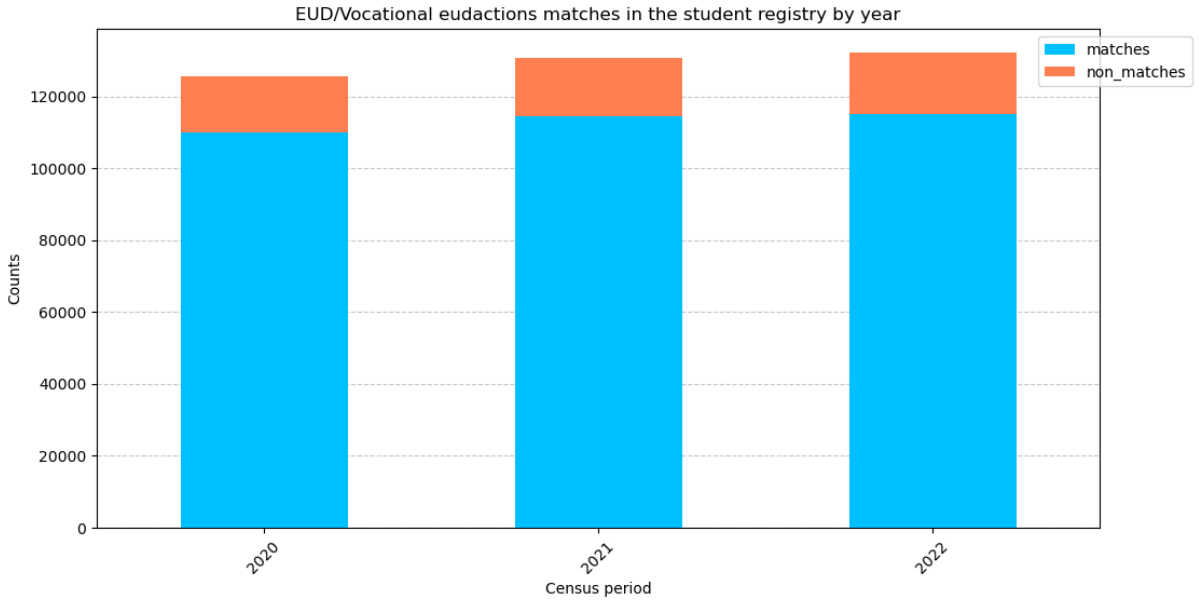
Strategic approach to quality measures for the matched educational data

DANMARKS STATISTIK

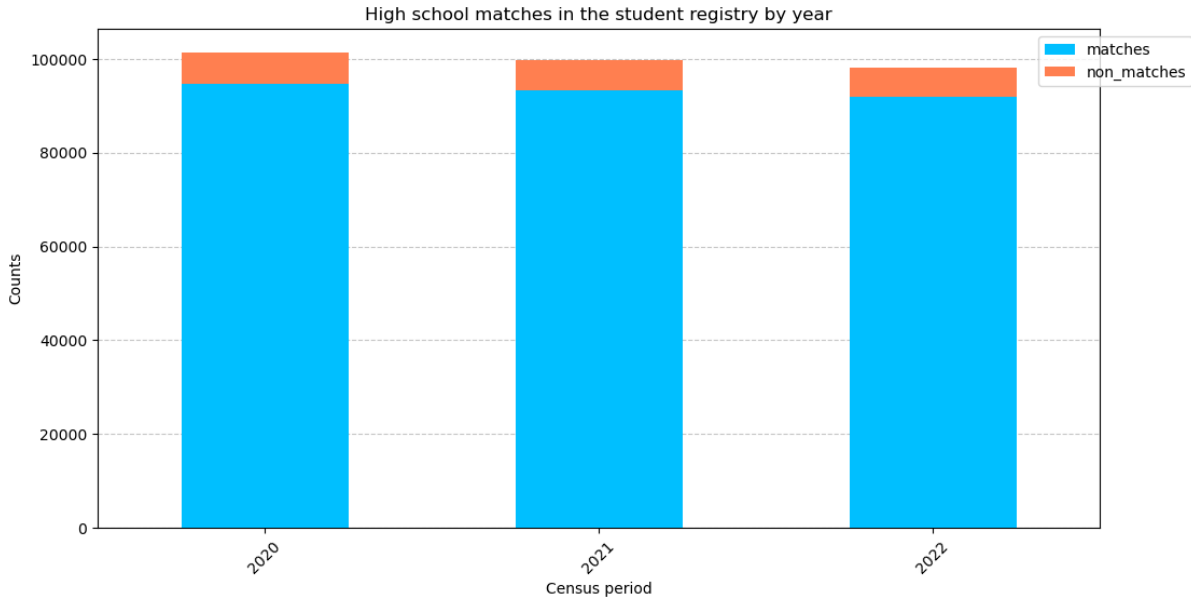
- What are the steps we take to validate the data?
  - Setup up pre validation checks such as format such as format and code classification checks
  - Automatically correct illogical data entries such as future dates
  - Automatically remove data entries not adhering to reference registers
  - Analyze the validity of the data population, and adherence to administrative rules
- Data analysis
  - Focused on the core of the MED: The student table
    - Matched the student population against the relevant student registry population
  - Further analysis
    - Teachers
    - Lessons
    - Classes.

Strategic approach to quality measures for the matched educational data

Student population matches from MED in the student registry for vocational educational programs



Student population matches from MED in the student registry for general secondary upper educational programs



Match criteria: Personal identification number, institution code, educational code and educational part

Student population matches from MED in the student registry for vocational educational programs

Match Rates of LER in the student registry by LER Criteria: cpr

Census period	Student entries LER	Matches in student registry	%
2020	125682	124627	99.16
2021	130859	129572	99.02
2022	132213	130730	98.88

Match Rates of LER in the student registry by LER Criteria: cpr, hovedskoleinstitution

Census period	Student entries LER	Matches in student registry	%
2020	125682	117198	93.25
2021	130859	121512	92.86
2022	132213	121787	92.11

Match Rates of LER in the student registry by LER Criteria: cpr id, hovedskoleinstitution, udd

Census period	Student entries LER	Matches in student registry	%
2020	125682	116232	92.48
2021	130859	120356	91.97
2022	132213	120405	91.07

Match Rates of LER in the student registry by LER Criteria: cpr id, hovedskoleinstitution, udd, udel

Census period	Student entries LER	Matches in student registry	%
2020	125682	109902	87.44
2021	130859	114348	87.38
2022	132213	115184	87.12

Student population matches from MED in the student registry for general secondary upper educational programs

DANMARKS STATISTIK

Match Rates of LER in the student registry by LER Criteria: cpr

Census period	Student entries LER	Matches in student registry	%
2020	101438	101122	99.69
2021	99800	99457	99.66
2022	98099	97659	99.55

Match Rates of LER in the student registry by LER Criteria: cpr, hovedskoleinstitution

Census period	Student entries LER	Matches in student registry	%
2020	101438	97108	95.73
2021	99800	95549	95.74
2022	98099	94003	95.82

Match Rates of LER in the student registry by LER Criteria: cpr id, hovedskoleinstitution, udd

Census period	Student entries LER	Matches in student registry	%
2020	101438	95721	94.36
2021	99800	94231	94.42
2022	98099	92754	94.55

Match Rates of LER in the student registry by LER Criteria: cpr id, hovedskoleinstitution, udd, udel

Census period	Student entries LER	Matches in student registry	%
2020	101438	94762	93.42
2021	99800	93390	93.58
2022	98099	91988	93.77

Strategic approach to quality measures for the matched educational data

DANMARKS STATISTIK

- Analysis of the teacher population:
  - No high quality source to validate against
  - Match against the population in the personnel tables delivered as part of a contribution to international statistics
  - Difficult to assess if populations are uniformly delimited

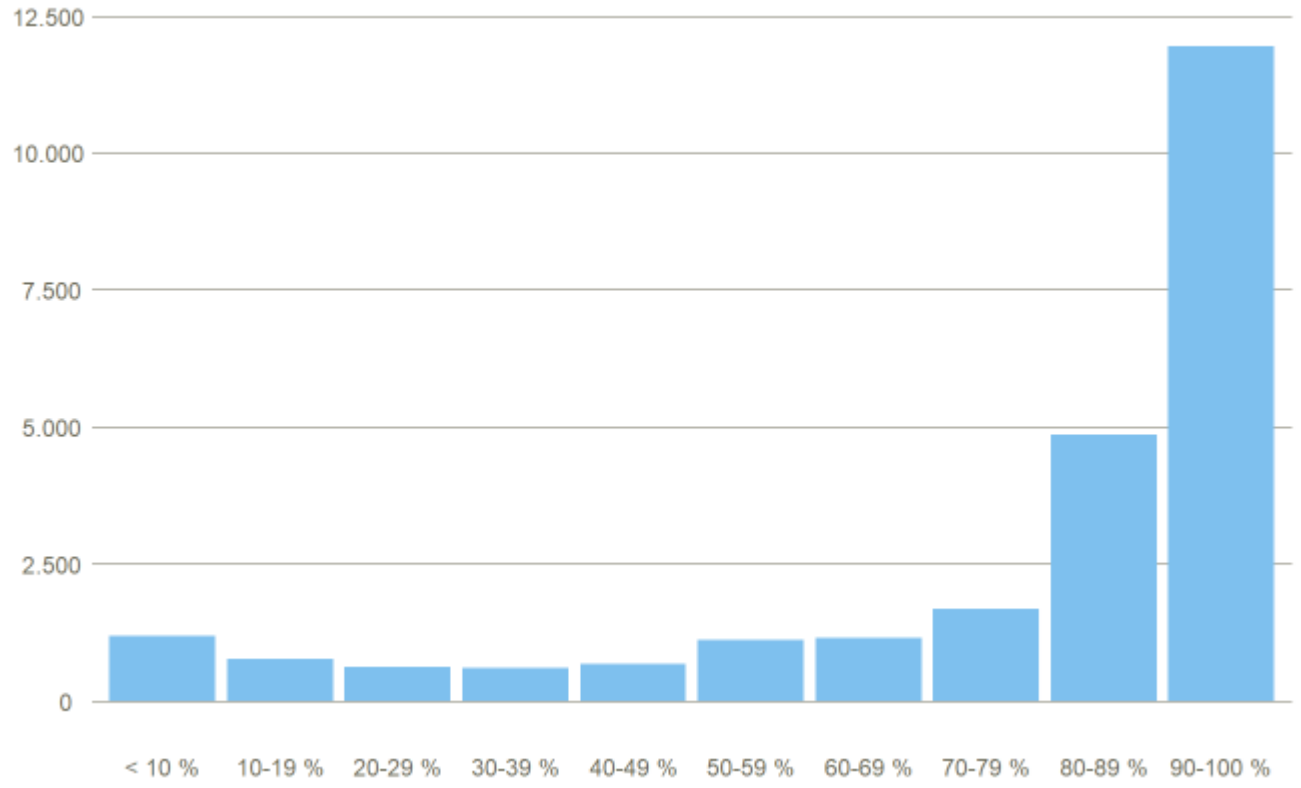
  Match percentage between 70 and 80% depending on the match criteria

- Analysis of the lessons for the vocational educational programs
  - No previous high quality source to validate against
  - Examine if the educational programs upholds the government mandated rules for minimum amount of lessons

Analysis of classes for the vocational educational programs

DANMARKS STATISTIK

Distribution of classes that adhere to the legal minimum of weekly teacher led lessons

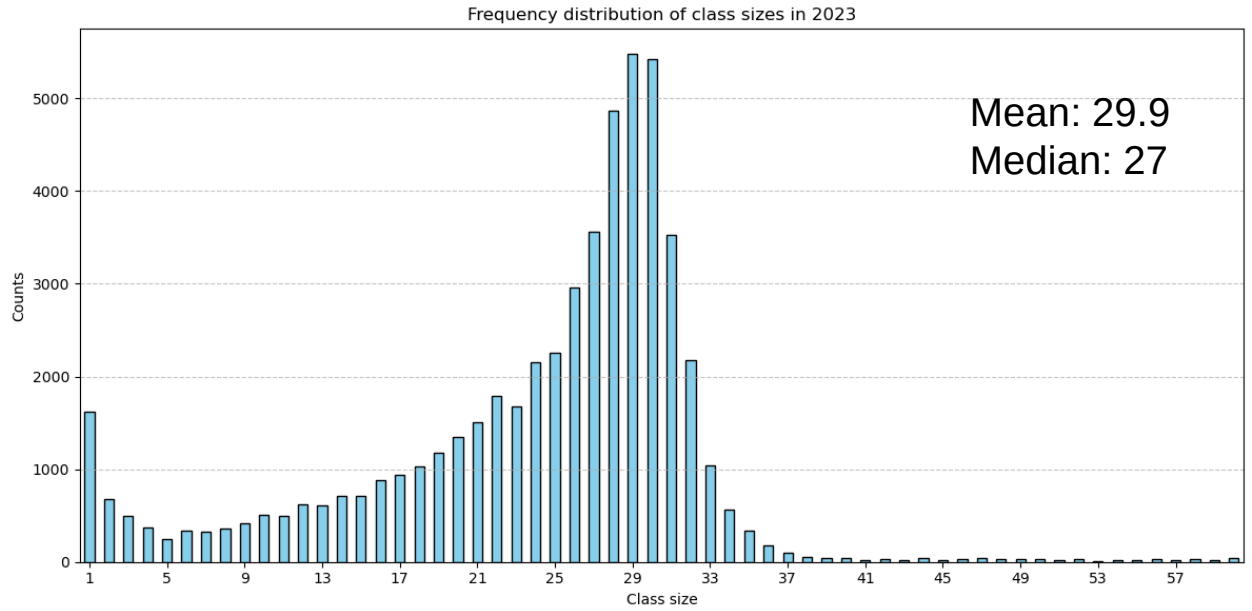


Strategic approach to quality measures for the matched educational data

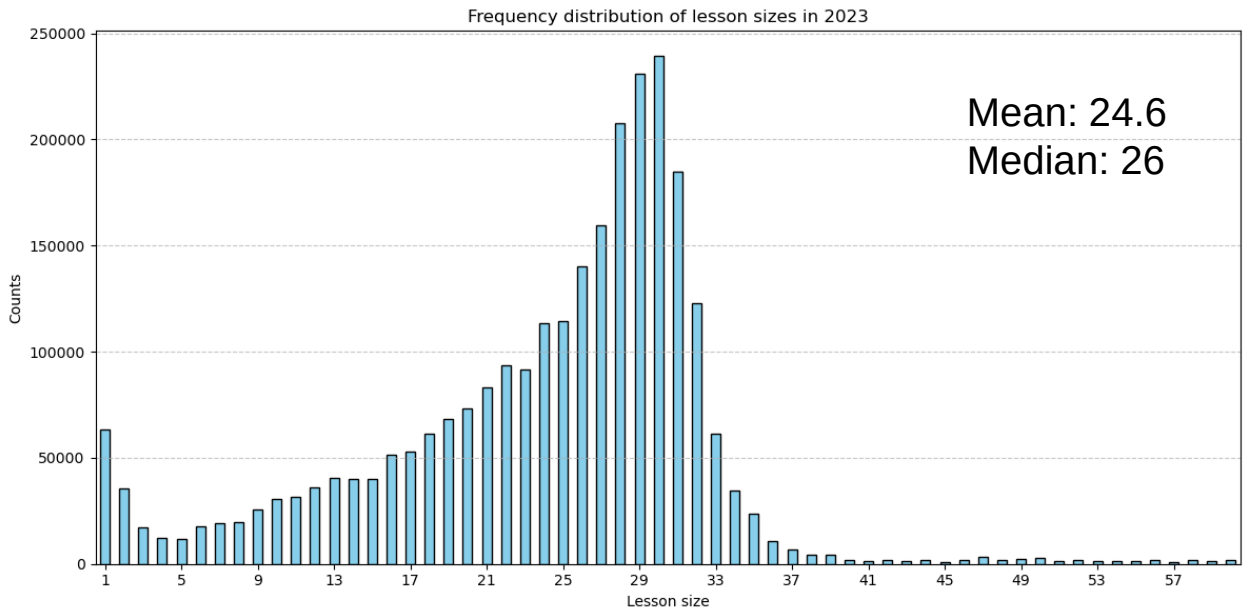
DANMARKS STATISTIK

- Analysis of class and lesson sizes for general upper secondary educations
  - Validate against the criteria for an average class size of 28.
  - Examine the frequency of the size distribution in classes and lesson for the school year 2023

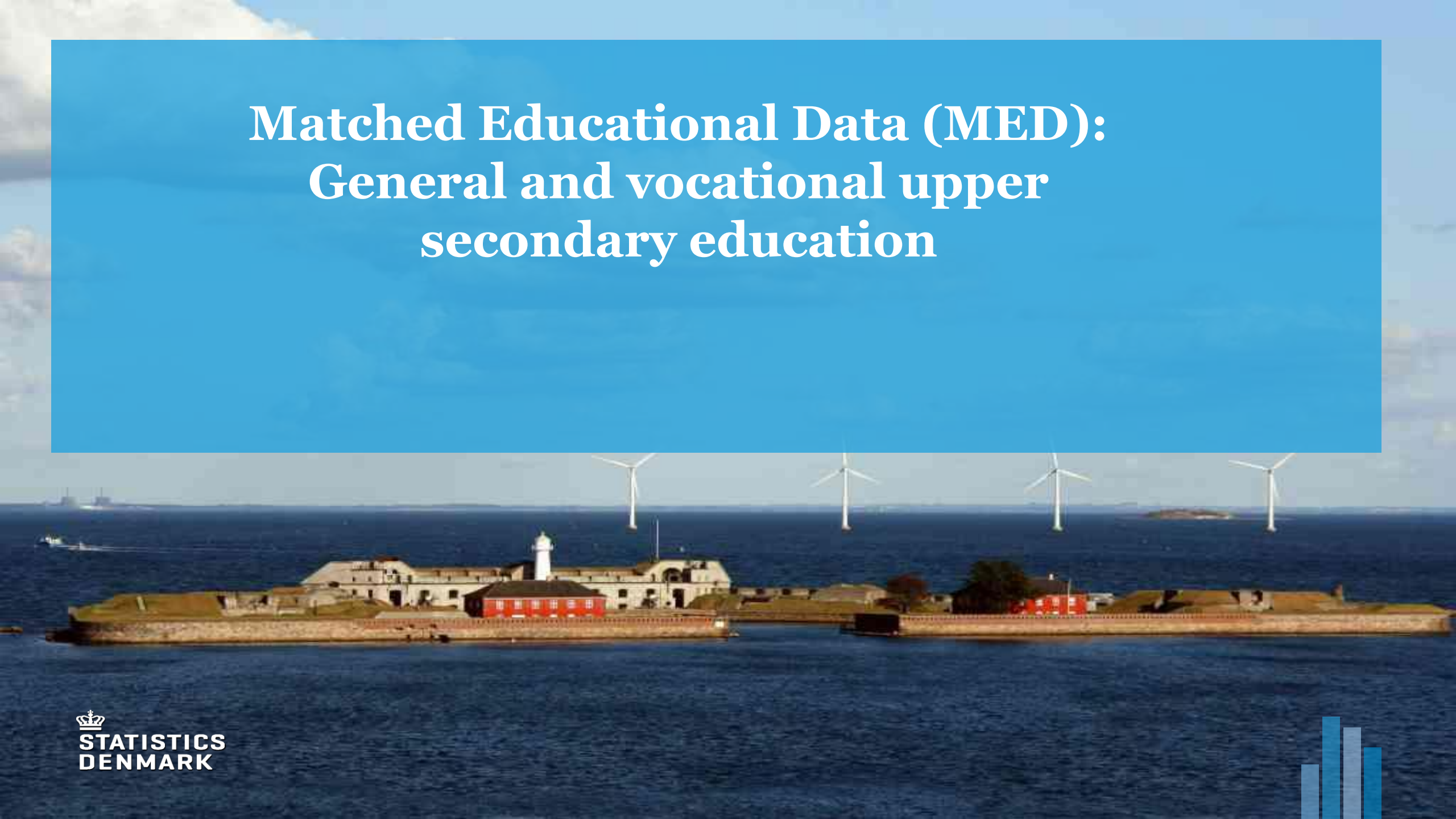
Distribution of class sizes for general upper secondary educations



Distribution of lesson sizes for general upper secondary educations



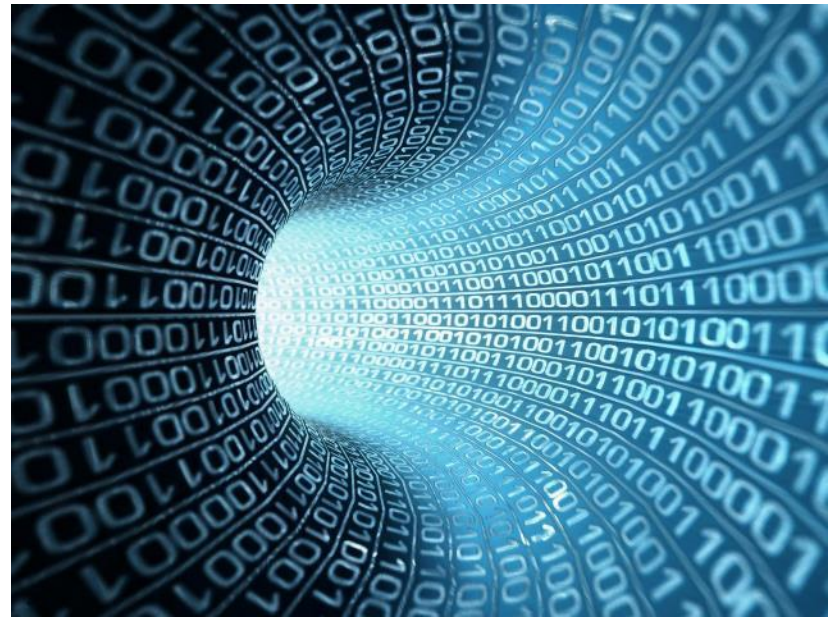
Matched Educational Data (MED): General and vocational upper secondary education



STATISTICS DENMARK

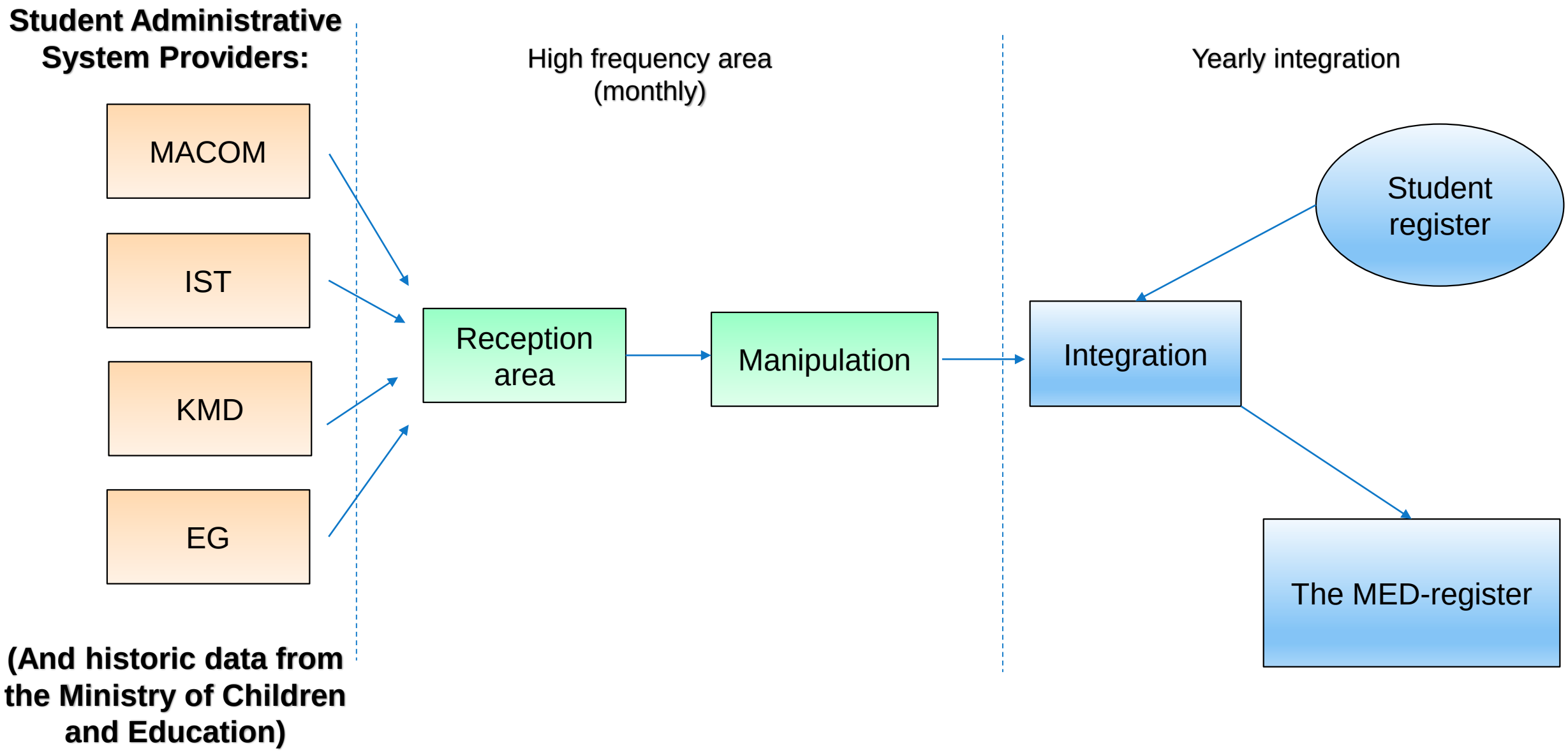
Data Collection And Scope

- Current coverage is data from MACOM and IST. Covering 89% of Danish Vocational Schools and 93% of Danish upper secondary schools. And students who were active in the period; 2020-2022.
- Work ongoing to use data from all system providers and cover all schools. And to extend coverage backwards in time. (in collaboration with the Ministry of Children and Education).

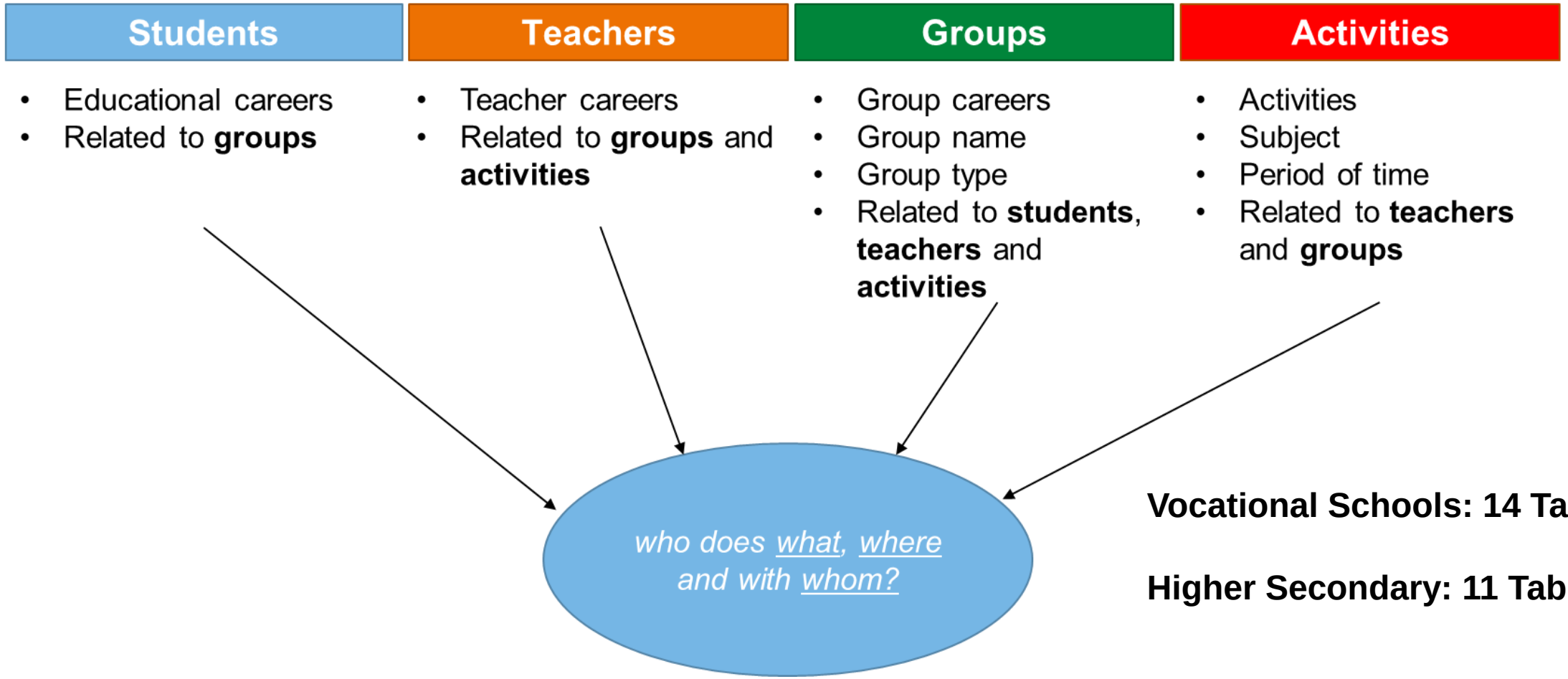


Collection and Production infrastructure

STATISTICS DENMARK



Model; Data separated into several categories of tables



Vocational Schools: 14 Tables

Higher Secondary: 11 Tables

How is this data different from previously available education data?

Some examples of joins and research questions

Example: General upper Secondary School day: lesson plan of student

Teacher specific effects

Institution number	Room	Subject Title	Level	Start_time	End_time	Teacher	Teacher Competency	Minutes of absence	Excused	Reason
515402	21	Math	B	08:10:00	09:30:00	ID 1	(Subject Competence)			
515402	21	German	B	09:45:00	11:05:00	ID 2	(Teaching Competence)	80	FALSE	Andet
515402	21	Danish	A	11:35:00	12:55:00	ID 3	(Teaching Competence)			
515402	21	Lecture		13:05:00	14:25:00					

Determinants or effects of absenteeism

Example: Vocational school day: Lesson plan of student

Timing or sequence effects

Institution number	Room	Subject Title	Level	Start_time	End_time	Teacher	Teacher Competency	Minutes of absence	Excused	Reason
461449	L2.21	Danish	C	08:15:00	09:45:00	ID 1	(Teaching Competence)	5	FALSE	Other
461449	L2.21	Pedagogy	Null	10:05:00	11:35:00	ID 2	(Teaching Competence)			
461449	L2.21	Pedagogy	Null	12:05:00	13:35:00	ID 3	(Subject Competence)	35	TRUE	Sick

Effects of illness on educational outcomes

Apprenticeship data + background variables

STATISTICS DENMARK

Class-specific Effects, or class-mate effects

Effect of Test scores on receiving apprenticeship

Student_ID	Class ID	Subject	Grade	Level	Apprenticeship	Apprenticeship business_ID	Minutes of absence	Excused	Absence Type
ID 1	Class 1	Math	4	C	No apprenticeship Agreement	NULL	5	FALSE	Andet fravaer
ID 2	Class 1	Math	7	C	Apprenticeship agreement	Business_ID 2			

Data Access

- The Research Service of Statistics Denmark makes microdataavailable through remote access to a secure research computer hosted by Statistics Denmark.
- A variety of access arrangements are available, from individual research project access, to institutional access.
- International researchers can gain access through collaboration with a Danish research institution.
- Not yet a standard Research Service product. You will have to contact your research service project manager for your customized data needs.
- Data will be available come May. We will be available for dialogue through the email provided earlier

Analysis questions investigated with MED;

Danish Economic Council: Teacher value added; effect of teacher productivity on grade results of students.

Copenhagen Business School – Fane Groes: Relationship between highly detailed (at lesson level) data and dropout rates (MED Component ongoing/underway)

DST Masters Student: Causes of dropping out
a) Effect student-teacher ratio
b) Effect gender composition of class / gender-corresponding teacher

STATISTICS DENMARK

Thank you for your time

Asger Bromose Langgaard,
Alexander Erik Friisnæs and
Christian Vittrup

Mail address for questions:
LER_Ungdomsuddannelser@dst.dk