

Transforming Danish Registries to the OMOP Common Data Model: use case on the Danish Colorectal Cancer Group (DCCG) Database

PRESENTER: **Andi Tsouchnika**

INTRO:

- Standardizing registry data to address multifactorial diseases like colorectal cancer.
- Enriching the standardization pipeline with additional layers of patient de-identification and quality control in order to comply with data authorities and produce a clinical grade CDM that can be used for direct patient interventions.

METHODS

Data: DCCG national clinical registry initiated at 2001, with 384 variables, ~77k patients.

Transformation process:

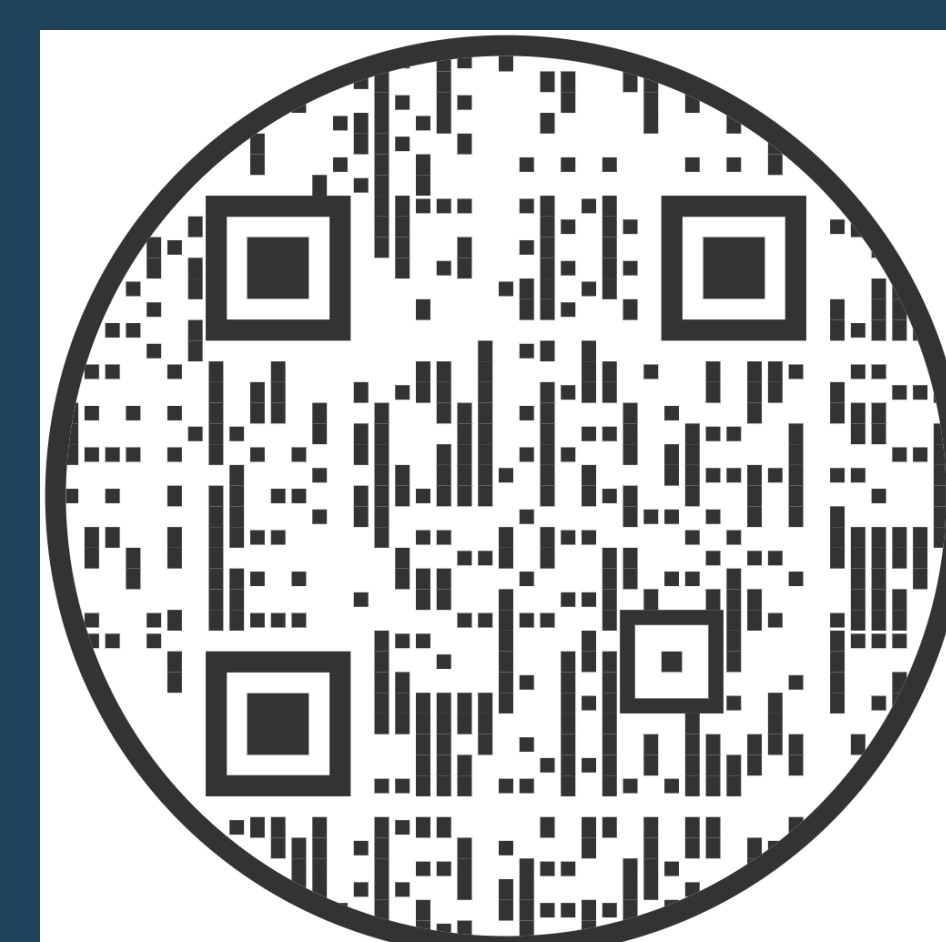
- Data Curation:** 1. invalid dates, 2. illogical values
- De-identification:** 1. hash direct ids, 2. remove infrequent events, 3. person date shift, 4. combine infrequent ages, 5. modify physiological measurements.
- OMOP ETL:** 1. *Concept mapping:* web application developed by edenceHealthNV, 2. *Structural mapping:* Rabbit-In-A-Hat and THEMIS conventions.

Evaluation of transformation:

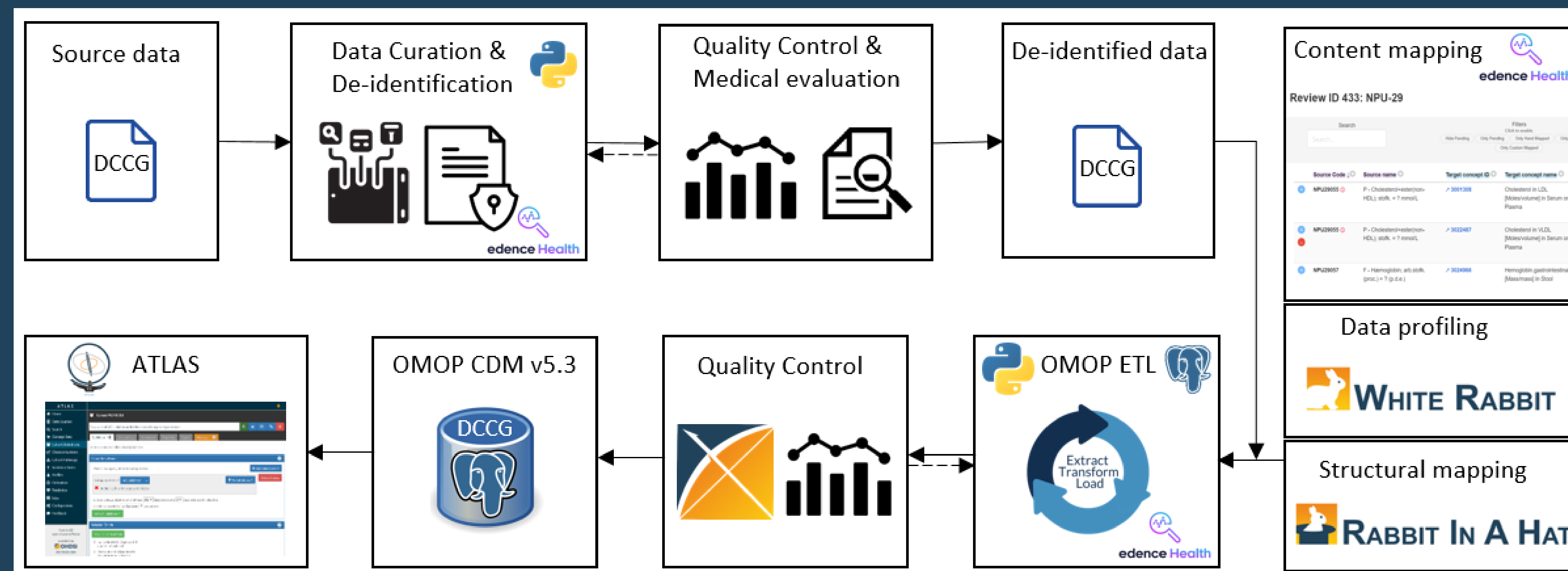
- DC&De-id:** 1. column-based logging with number of affected records from each rule, 2. statistical comparison and medical evaluation.
- OMOP ETL:** 1. statistical comparison, 2. column-based logging, 3. random patient spot-check, 4. Data Quality Dashboard.
- Overall assessment:** descriptive statistics of 16 selected variables between source, de-identified and CDM data and association with 30-day post-operative death using logistic regression.

RESULTS

- Mapping coverage of 99,12% (85% standard and 15% custom mappings).
- 76.849 rows and 317 variables to 9.668.672 records in OMOP CDM without losing any patients.
- 224,869 data points were lost because of missing dates.
- DC&De-id steps did not affect the general distributions of the data.
- Overall assessment showed no significant difference between source, de-id and CDM data.



De-identification and Quality Control steps to comply with data authorities and create a clinical grade CDM



TABlename	COUNT	PATIENTS
observation	5,124,260	76,849
condition_occurrence	1,169,054	76,849
measurement	1,079,371	76,847
condition_era	878,790	76,849
procedure_occurrence	750,378	74,557
visit_occurrence	308,406	76,849
specimen	139,465	40,479
observation_period	76,849	76,849
person	76,849	76,849
death	42,620	42,620
drug_era	8,784	4,953
drug_exposure	8,784	4,953
location	4,678	-
care_site	384	-
device_exposure	0	0
note	0	0
dose_era	0	0
cost	0	-
payer_plan_period	0	0
visit_detail	0	0
provider	0	-

TARGET VOCABULARY ID	COUNT
SNOMED	2768
CSS (custom)	479
LOINC	62
None	30
OMOP Extension	28
Nebraska Lexicon	15
CPT4	7
RxNorm	7
HemOnc	2
ICD03	1
NAACCR	1

Andi Tsouchnika, Maliha Mashkooor, Andreas Weinberger Rosen, Eldar Allakhverdiiev, Jared Houghtaling, Freija Descamps, Mikail Gögenur, Viviane Annabelle Lin, Johan Stub Rønø Clausen, Karoline Bendix Bräuner, Julie Sparholt Walbech, Soham Ravindra Shinde, Ismail Gögenur

