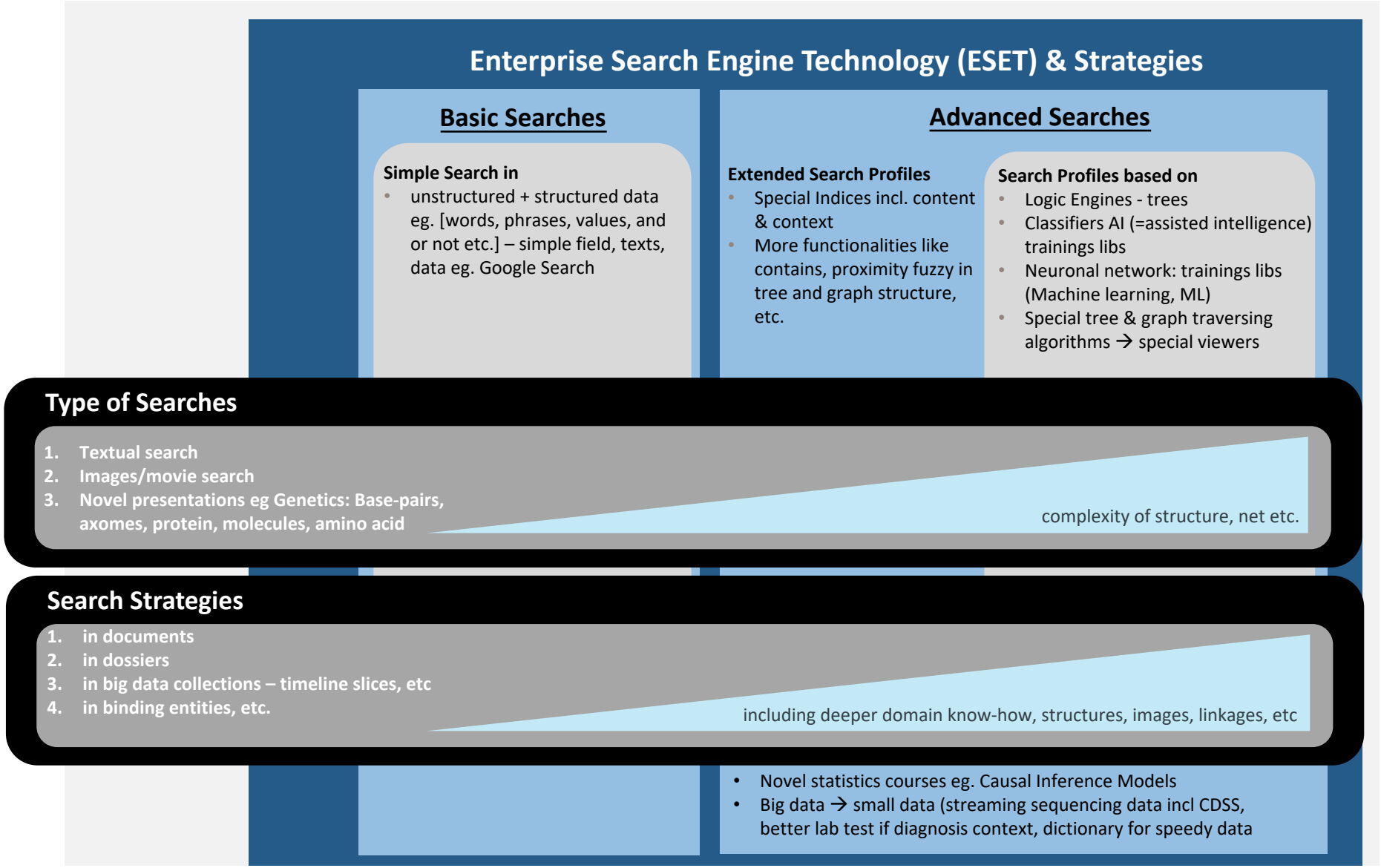
Introduction Search Profiles

ITERATA EXPLAINS

A CRITICAL APPROACH TOWARDS THE SUBJECT

Advanced Searches from BIG Data to SMALL Data



Cohort Identification

prior the search

- Concrete research question (PICO)
  - H0
  - H1
- Definition exclusion/inclusion criteria (age, time range, gender, diagnosis, side effects, medication, therapies…)

→ **Set of searchable attributes**

- Creating search profile

search request

- Running created search profile (e.g. 1'000)
- Adjusting search profile (add/remove attributed)
- Rerunning search profile (e.g. 800)

→ **feasible # of patients in cohort**

Cohort Validation

Validation

- Manual validation of patients based on inclusion/exclusion criteria
- Reasoning of inclusion/exclusion of patients (yes/no)

→ **Final cohort extracted (e.g. 300)**

Advanced validation

- Automated validation based on predefined classifier sets (machine learning methods)

Cohort Structuring

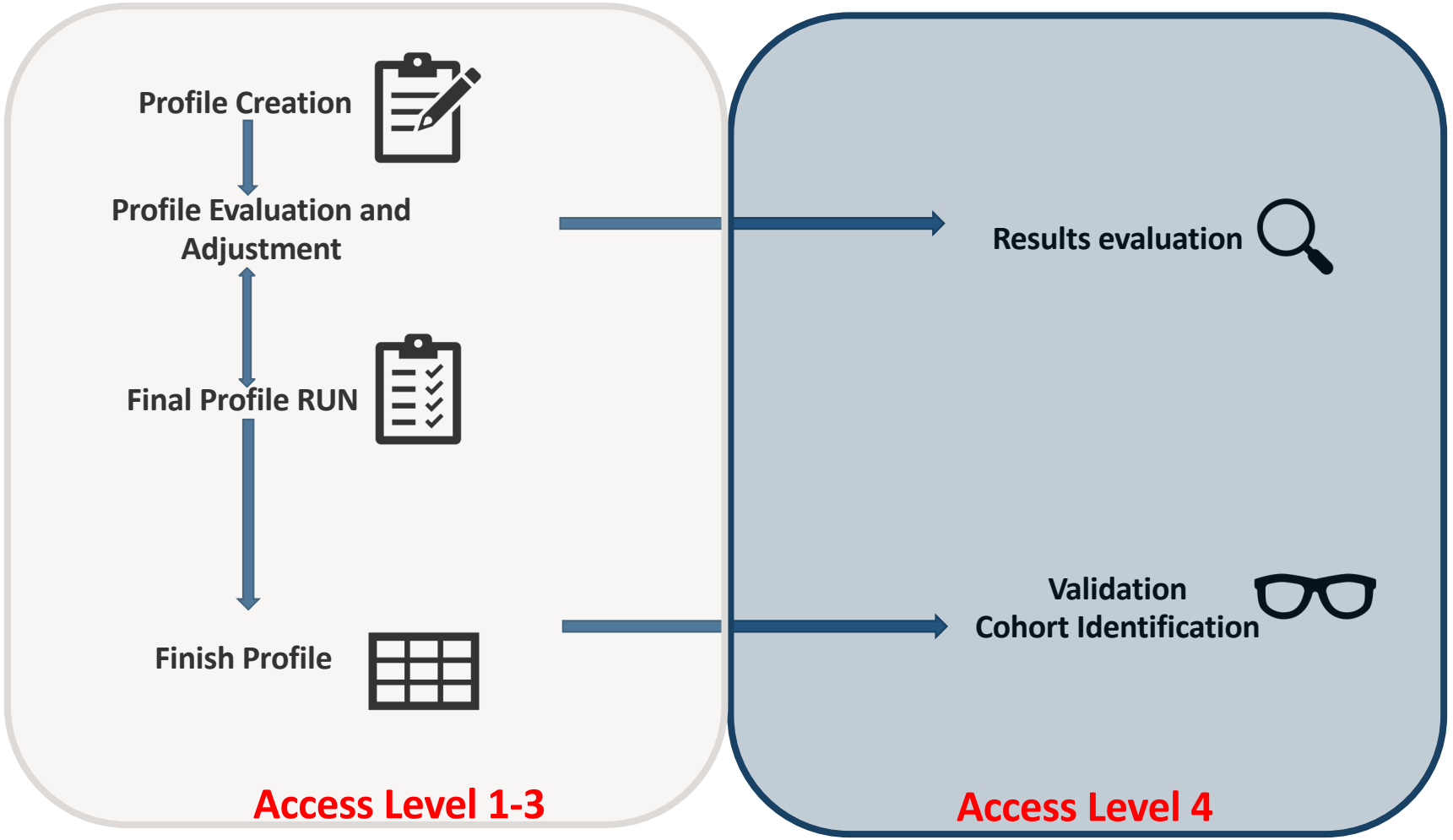
Structuring

- Search results structured in table form
  - Rows: patients
  - Column: search attributes

Statistics

- How to evaluate the search results

→ Statistic curriculum


Profile Creation
Profile Evaluation and Adjustment
Final Profile RUN
Finish Profile
Results evaluation
Validation
Cohort Identification
Access Level 1-3
Access Level 4

Cohort Identification

Prior the search - PICO

Development of question of interest that is as precisely as possible

- Based on this, extraction of keywords for search profile
- Determine hierarchical order of keywords

P: patient/population/ problem (age, gender, diagnosis, side effects)

I: intervention (treatment, medication, therapies, surgery, predictors)

C: comparison (alternative treatments, control group – for interventional studies)

O: outcome (treatment aim, outcome measures)

Additional:

T: Time (defined time frame)

S: study design (RCP, observational)

Search Profile Creation

	colume A	colume B	colume C	colume D
row 1	Vorhofflimmern			
row 2		Herzinsuffizienz		
row 3			Schlaganfall, TIA, Zerebrale transitorische Ischämie	
row 4				Hämatemesis
row 5		Herzinsuffizienz		
row 6			Hypertonie, Bluthochdruck~	Hämatemesis
row 7		Herzinsuffizienz		
row 8			Diabetes mellitus ~	
row 9			patientBirthdate=1946<>1955	
row 10		Schlaganfall		
row 11			Hämatemesis ~	
row 12	Biopsie			
row 13		Transplantation		
row 14			Nieren	
row 15		!Diabetes melitus		
row 16		Lebertransplantation		
row 17		*biopsie		
row 18		Lungen*		

- **AND** eg. («Vorhofflimmern AND Herzinsuffizienz») [same page]
- **AND** & **OR** eg. («Vorhofflimmern AND Herzinsuffizienz AND (Schlaganfall OR TIA OR Zerebrale transitorische Ischämie) AND Häatemesis») [same page]
- **Proximity search** eg. «Diabetes melitus~»
- **OR**
  eg. («Vorhofflimmern AND Herzinsuffizienz AND Diabetes mellitus)»)
  OR («Vorhofflimmern AND Herzinsuffizienz OR patientBirthdate=1946<>1955»)
- **Fuzzy + Factor** eg. «Hämatemesis~»
- **NOT** eg. «Diabetes melitus»
- **End-with** eg. *biopsie = Nierenbiopsie, Leberbiopsie, etc.
- **Start-with** eg. Lungen* = Lungenleiden, Lungenkrankheit, etc.
- **Substring** combination eg. *lungen* =Herzlungenmaschine

Semantic for Excel Tree check in:

- **AND**: An AND Linkage is represented in a row (horizontal merges). exact phrases between “”Each term is in the same row in the 1..n columes
- **OR**: An OR Linkage is represented in combination of more than one row (vertical merges)
- **OR** in cell: [item, item, … , item] = [item OR item OR …. OR item]
- **NOT**: is represented as **!**[item or phrases]
- **End-with**: ***[item]; **Start-with**: [item]***; **Substring: ***[item, phrases]***

All searches e.g. cell items or phrases **« »** will be executed as an exact term search, but with ~ at the end a **Proximity search** will be performed: [item item]~

- **Fuzzy**: Just per item could be weighted with a fuzzy factor, e.g. spelling mistakes “Aschpirin”, “Vorhoflimern”

Way of Searches - Result Set

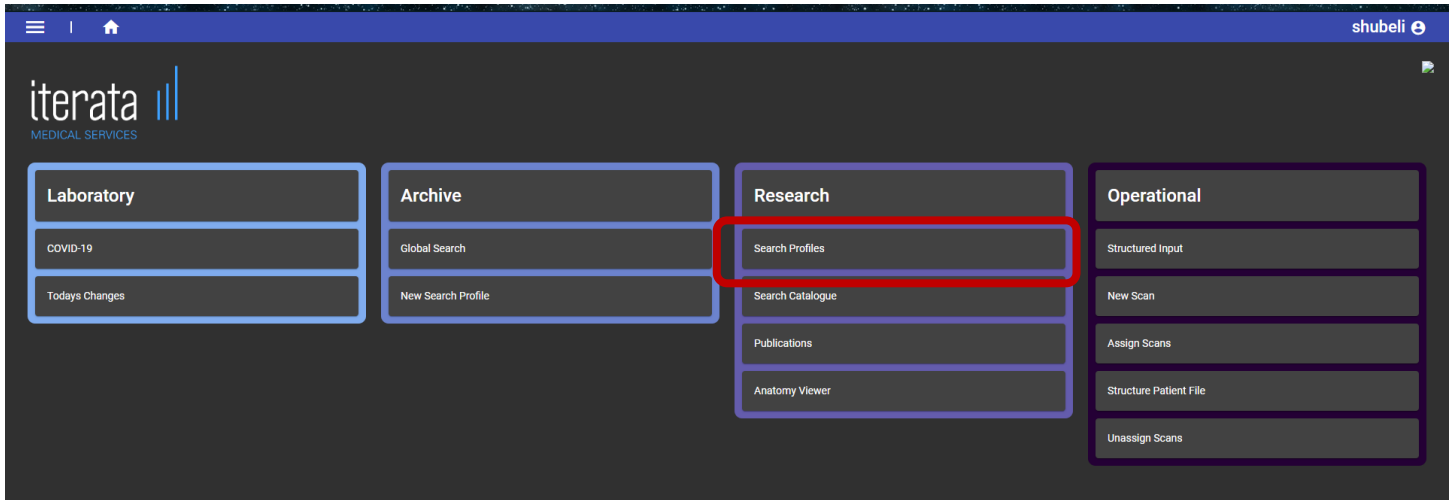
Way of Searches

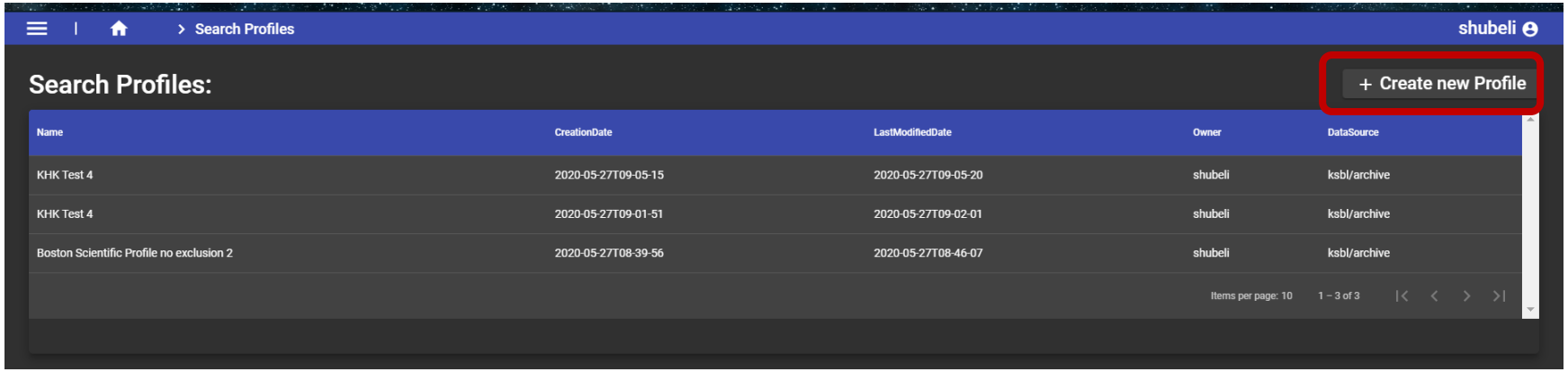
1. **In one report (all criteria's) - possible**
2. **In one patient dossier (throughout all concatinated reports) – possible for dedicated searches**
3. **All in Dictionary – referenced to all pat and docs – coming soon (graph database)**

Result Set

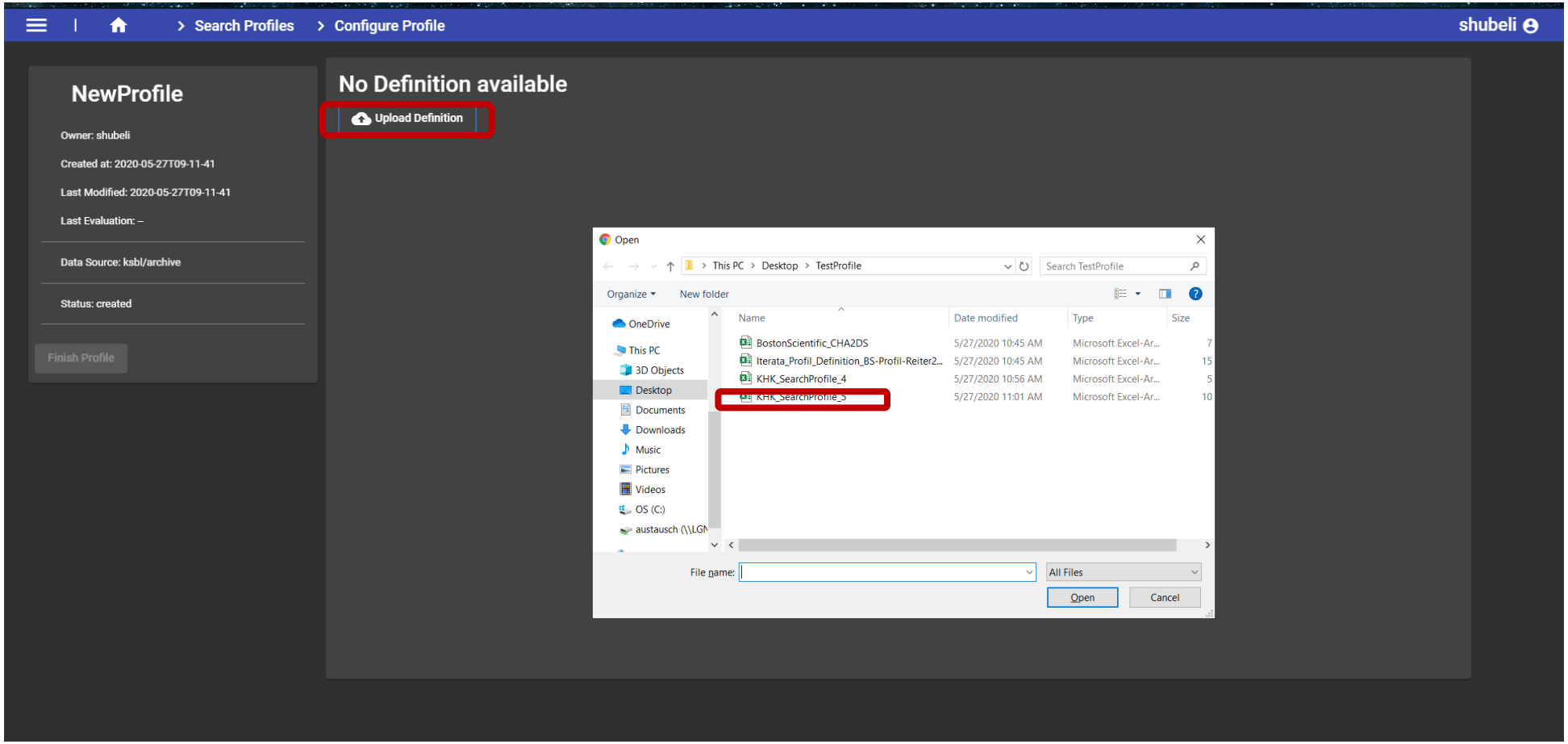
- **# Documents and #Patients per Combination (row)**
- **Over selected Row's → Freeze of potential cohort**

Check in Profile

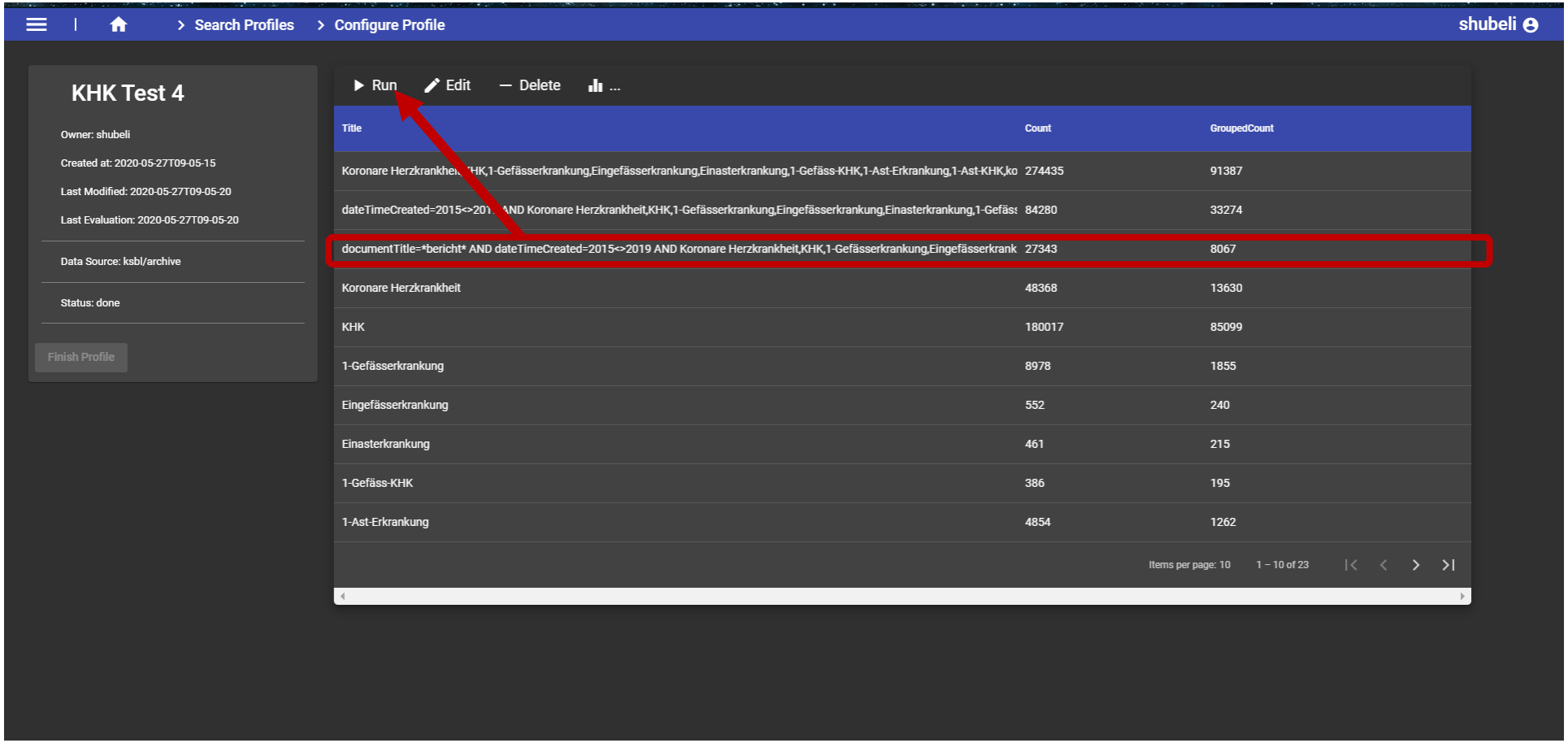




Check in Profile – Select Profile Definition

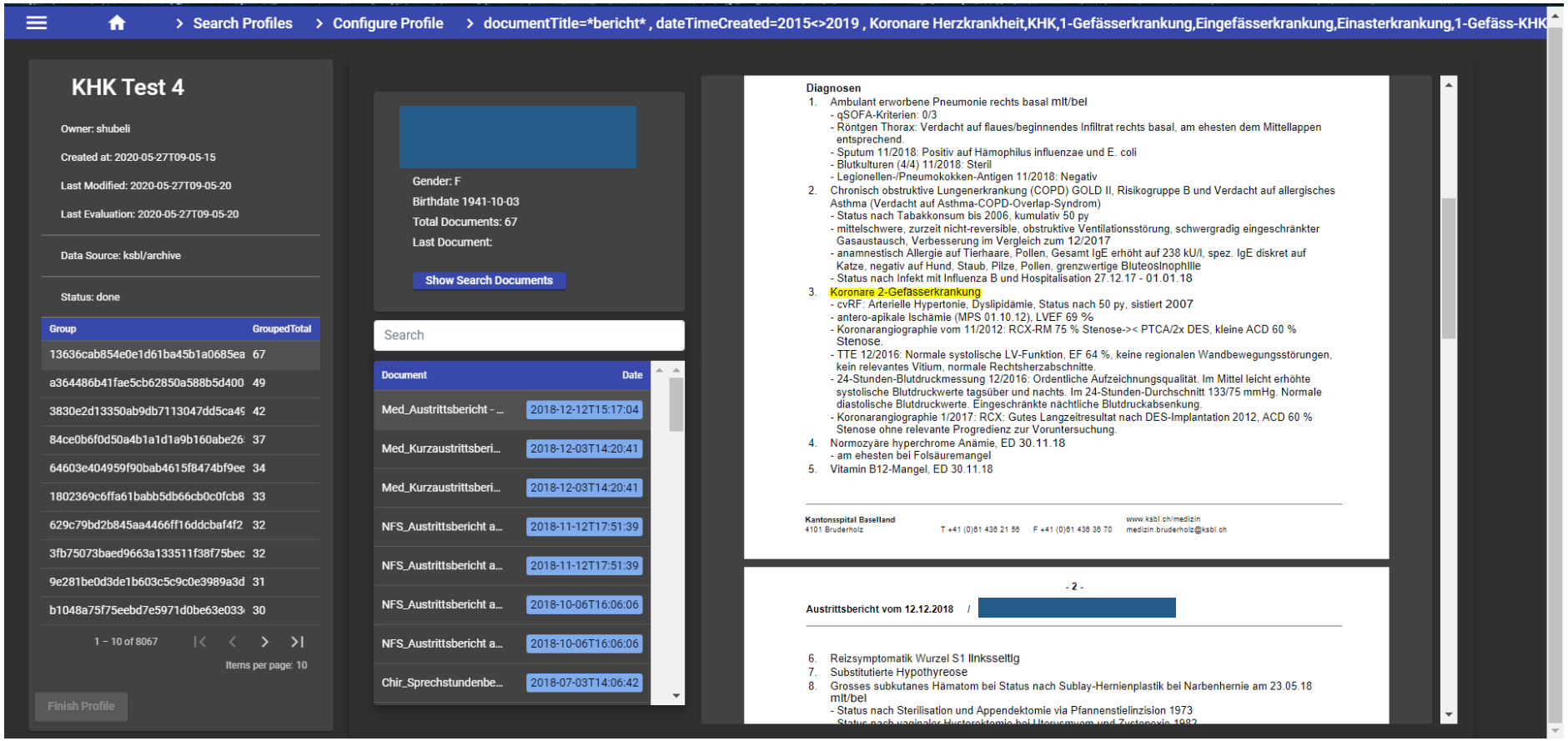


RUN Profile → Result per each combination of items or phrases



Cohort Validation

Validate Result of Profile on Level 4



Cohort Structuring

What's next

Correlation ≠ Causation

AND

Data ≠ Knowledge

→ **Data validation, critically question findings, put results into domain-specific context!**

5 principles to build on to

- **Human role**: stand on the power of computers, don't compete with them (**Use-cases, research question**)
- **Mainstream**: Establish a mainstream, core computational thinking curriculum for all (**Curriculum**)
- **Realistic**: Use the real-world as your guide for what to learn: concepts, strategies and toolsets (**Data quality, evaluation of data**)
- **Toolset**: Priorities breadth of computational applications over the details of their implementation (**Search strategies, ML (Classifier), LogicEngine**)
- **Urgency**: Implement computational thinking education – society's key preparation for the AI age.

Key essence: Applying the available tools considering scientific methods, critically question common methods and don't use them just because it has always been done that way.