

# Avatud andmed

**Jaak Vilo**

[cs.ut.ee](http://cs.ut.ee)

**Tartu Ülikool**

**Arvutiteaduse instituut**

**Andmeteanduse õppetool**



# Avaandmete foorum 2023

Online event

About Discussion

## Details

# Tallinnas, L'Embitu hotellis kell 9.30-16.00.

🕒 6 hr 30 min

👤 177 people responded



👤 Event by Majandus- ja Kommunikatsiooniministeerium

🌐 Public · Anyone on or off Facebook

🔔 Kuuendat korda toimuv Avaandmete foorum toob taas kokku avaandmete valdkonna huvilised. Sellel aastal arutame avaandmete rolli riigijuhtimises, avaandmete ümber kogukonna loomist ja avaandmete kasutamise võimalusi.

Oodata on mitmeid esinejaid, kes jagavad kogemusi ja ideid, kuidas nemad oma töös avaandmeid kasutavad ning millist väärtust need loovad.

Päevakavas on välkettekanded ja paneeldiskussioonid, kus räägime näiteks:

- ◆ Avaandmete valdkonna hetkeseisust ja tulevikust
- ◆ Avaandmete rollist riigijuhtimises
- ◆ Kogukonna loomisest avaandmete ümber
- ◆ Avaandmete kasutamisest panganduses, telekommunikatsioon, hariduses ja muudes valdkondades
- ◆ Kohalike omavalitsuste avaandmete projektidest

Osalema on oodatud kõik andmehuvilised!

Avaandmete foorum toimub nii füüsiliselt kui ka virtuaalselt.

REGISTREERIMINE toimub kolmel viisil:

📄 FÜÜSILISEKS osalemiseks

- ◆ RTIPI (riigitöötaja iseteenindusportaal) ligipääsuõiguse olemasolul palume registreeruda siin: <https://tinyurl.com/zc8mtpdf>
- ◆ RTIPI ligipääsuõiguse mitte omamise korral palume registreeruda siin (vali üritus nimega „Avaandmete foorum“): <https://tinyurl.com/yrrurunh>

Registreerumine füüsiliseks osalemiseks on avatud kuni 29.09.23, üritus toimub Tallinnas, kohtade arv piiratud.

- ◆ RTIPI (riigitöötaja iseteenindusportaal) ligipääsuõiguse olemasolul palume registreeruda siin: <https://tinyurl.com/zc8mtpdf>
- ◆ RTIPI ligipääsuõiguse mitte omamise korral palume registreeruda siin (vali üritus nimega „Avaandmete foorum“): <https://tinyurl.com/yrrurunh>

Registreerumine füüsiliseks osalemiseks on avatud kuni 29.09.23, üritus toimub Tallinnas, kohtade arv piiratud.

📄 VIRTUAALSEKS osalemiseks registreeri siin: <https://tinyurl.com/2p9ckvu8>

Link live-ettekandele edastatakse registreerunutele e-posti teel hiljemalt ürituse toimumispäeval 12.10.2023.

KAVA:

- ◆ 09.30-09.40 Avaandmete foorumi 2023 avamine
- ◆ 09.40-10.10 Eesti avaandmete valdkonna hetkeseis - Ott Velsberg (MKM)
- ◆ 10.10-11.10 I välkettekannete sessioon

- Hannes Juuse (Eesti Pank)

- Rivo Roosileht (Telia)

- Veiko Lember (TalTech)

- TBA

- ◆ 11.10-11.40 Kohvipaus
- ◆ 11.40-12.30 Paneeldiskussioon „Avaandmete poliitilise rolli mõtestamine“

Osalejad: Dmitri Burnašev (Riigikantselei), Rivo Roosileht (Telia), TBA

Modereerib: Liina Osila (Statistikaamet)

- ◆ 12.30-13.30 Lõuna
- ◆ 13.30-14.30 II välkettekannete sessioon

- Hanna-Greta Veersalu (Tallinna Strateegiakeskus)

- Jaak Vilo (Tartu Ülikool)

- TBA

- ◆ 14.30-14.50 Kohvipaus
- ◆ 14.50-15.40 Paneeldiskussioon „Kogukonna loomine avaandmete ümber“

Osalejad: Hille Hinsberg (MTÜ Open Knowledge Estonia), Maarja-Leena Saar (SA Eesti Koostöö Kogu), TBA

Modereerib: Sofia Paes (MKM)

- ◆ 15.40-15.50 Päevakokkuvõtte

See less

# Ettakande teemad:

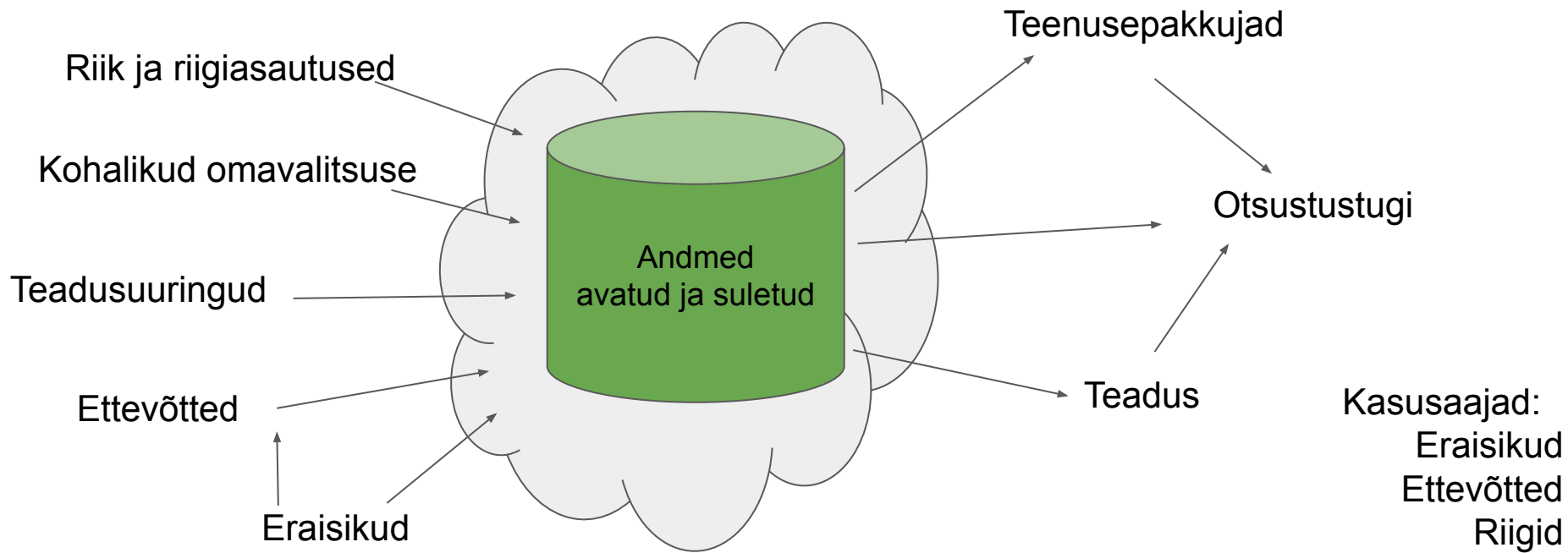
## 1. Avatud andmed on olulised

- a. Andmed kallid toota, neid tuleb taaskasutada
- b. Suur silmapaaride arv annab kasulikke tööriistu ja tulemusi

## 2. FAIR

## 3. Terviseandmed

- a. COVID-19 andmed ja rahvusvaheline koostöö
- b. Andmete teisene kasutamine!
- c. Sensitiivsuse aspektid
- d. OHDSI OMOP CDM - avatud andmemudelid ja tööriistad, koostöö



# Andmed tekivad mängimise käigus:



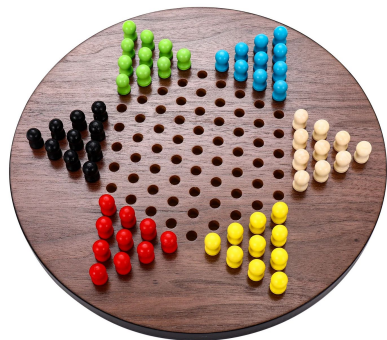
1950



1960; **1994** ; 2007



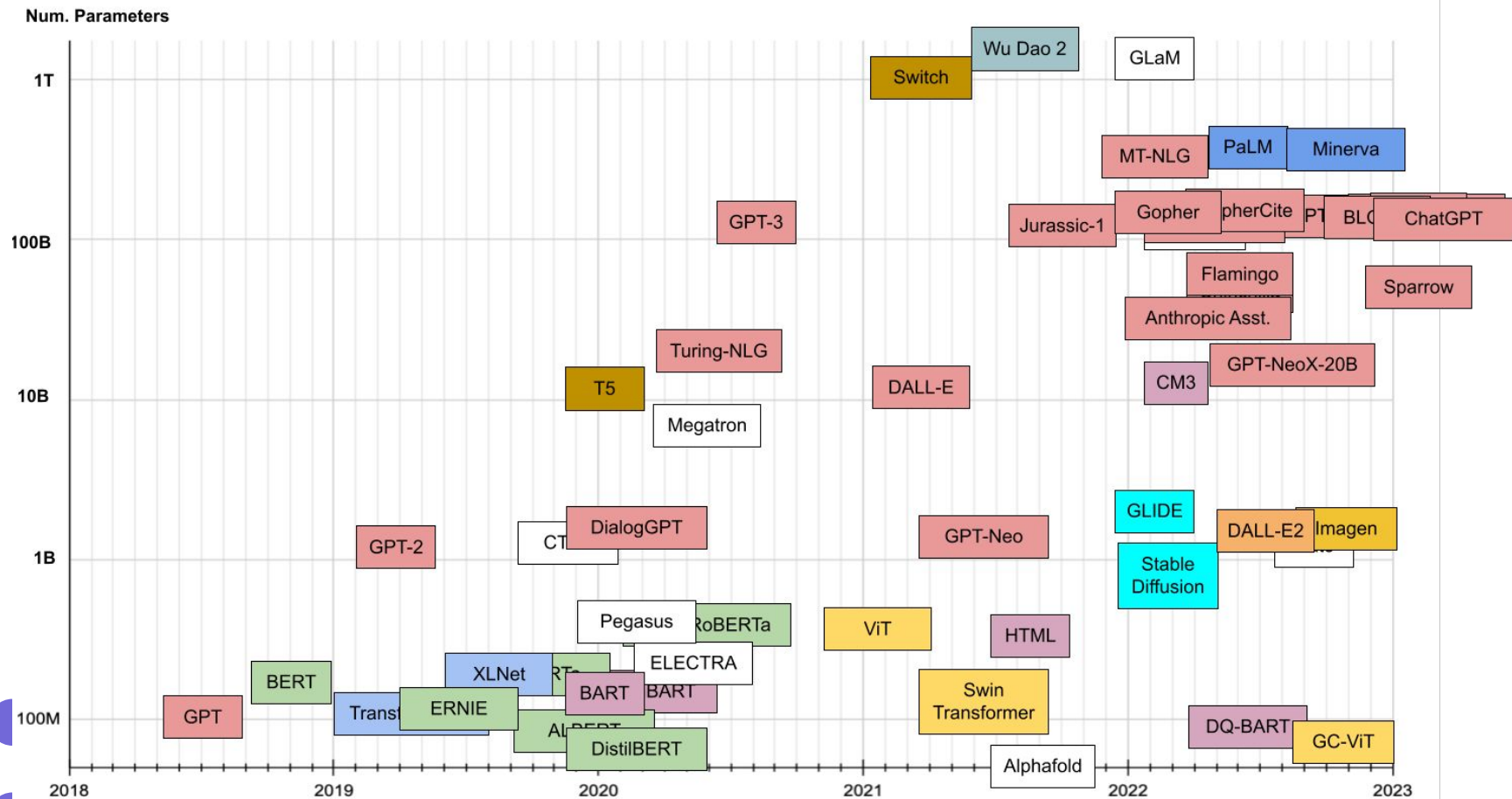
**1997** (IBM),



**2016**

**2017** - inimene ei pidanud ise ette andma ühtegi (go) mängu strateegiat

# 2. LLMs for Estonian language (Kairit Sirts, TÜ)

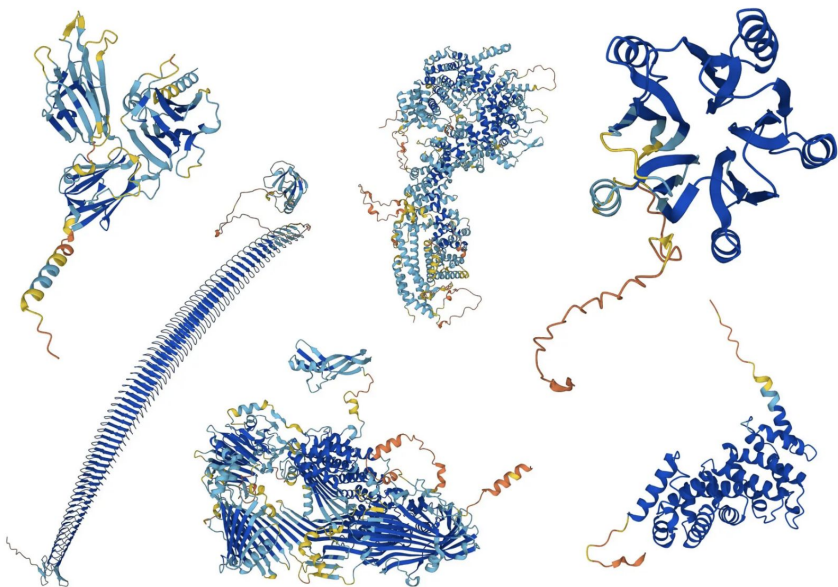


Amatriain, X. (2023). Transformer models: an introduction and catalog.

## *A.I. Predicts the Shape of Nearly Every Protein Known to Science*

DeepMind has expanded its database of microscopic biological mechanisms, hoping to accelerate research into all living things.

Share full article



Six protein shapes predicted by AlphaFold, an artificial intelligence technology under Google DeepMind.

# The New York Times

## Concepts

60+ years of research

Collecting data:

- Proteins, structures
- Protein sequence
- DNA structure 1953
- DNA sequence

## Competitions!



# 15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction



- Menu
- [Home](#)
- [PC Login](#)
- [PC Registration](#)
- ▼ [CASP Experiments](#)
  - [CASP15 \(2022\)](#)
  - [CASP14 \(2020\)](#)
  - [CASP Commons \(COVID-19, 2020\)](#)
  - [CASP13 \(2018\)](#)
  - [CASP12 \(2016\)](#)
  - [CASP11 \(2014\)](#)
  - [CASP10 \(2012\)](#)
  - [CASP9 \(2010\)](#)
  - [CASP8 \(2008\)](#)
  - [CASP7 \(2006\)](#)
  - [CASP6 \(2004\)](#)
  - [CASP5 \(2002\)](#)
  - [CASP4 \(2000\)](#)
  - [CASP3 \(1998\)](#)
  - [CASP2 \(1996\)](#)
  - [CASP1 \(1994\)](#)
- ▶ [Initiatives](#)
- ▶ [Data Archive](#)
- [Proceedings](#)
- [CASP Measures](#)
- [Feedback](#)
- [Assessors](#)
- [People](#)
- [Community Resources](#)
- [Job Fair](#)

## CASP15

CASP provides an independent mechanism for the assessment of methods of protein structure modeling. From May through August 2022, CASP organizers have been posting on this website sequences of unknown protein structures for modeling. Protein models have been collected from May through mid-September, and evaluated as the experimental coordinates become available. In the summer and fall, the tens of thousands of models submitted by approximately 100 research groups worldwide are processed and evaluated. Independent assessors in each of the prediction categories bring independent insight into their assessment. Tools for viewing, comparison, and analysis of submitted models are available from this website.

Targets	Predictors	Conference	Results
<a href="#">Target List</a>	<a href="#">Groups Info</a>	<a href="#">Abstracts</a>	<a href="#">REGULAR (domains)</a>
<a href="#">Domain Definition</a>		<a href="#">Program</a>	<a href="#">MULTIMERS</a>
		<a href="#">Presentations</a>	<a href="#">INTER-DOMAIN</a>
		<a href="#">Recordings</a>	<a href="#">EMA (ACCURACY)</a>
			<a href="#">RNA</a>
			<a href="#">LIGAND</a>
			<a href="#">Parseable Data</a>
			<a href="#">Rankings: Regular targets</a>
			<a href="#">Rankings: Multimeric targets</a>
			<a href="#">Rankings: Inter-domain prediction</a>
			<a href="#">Rankings: EMA prediction</a>
			<a href="#">Rankings: RNA prediction</a>
			<a href="#">Rankings: Ligand prediction</a>
<a href="#">CASP15 in numbers</a>			

## Detailed description of the experiment

CASP (Critical Assessment of Structure Prediction) is a community wide experiment to determine and advance the state of the art in modeling protein structure from amino acid sequence. Every two years, participants are invited to submit models for a set of proteins for which the experimental structures are not yet public. In the latest CASP round, CASP15, nearly 100 groups from around the world submitted more than 53,000 models on 127 modeling targets in 5 prediction categories. Independent assessors then compare the models with experiment. Assessments and results are published in a special issue of the journal PROTEINS ([check the latest CASP14 issue here](#)).

- [Goals](#)
- [Categories](#)
- [Timetable](#)
- [Registration](#)
- [Targets](#)
- [Format](#)
- [Assessment](#)
- [Results](#)
- [Conference](#)
- [Organizers](#)

## Message Board

**CASP SIG on Modeling Ensembles and Alternative Conformations - Wah Chiu and Martin Blackledge - Wednesday October 4, 2023**

[The next meeting of CASP SIG on Modeling Ensembles and Alternative Conformations will be held by zoom on Wednesday October 4, 2023 at 11:00 AM EDT \(New York time\). We will have 2 presentations, ab ...](#)

**September 6 - CASP SIG on Modeling Ensembles and Alternative Conformations**

[The next meeting of CASP SIG on Modeling Ensembles and Alternative Conformations will be held by zoom on Wednesday September 6th, 2023 at 11:00 AM EDT \(New York time\). Agenda for Weds September 6t ...](#)

**The third meeting of the CASP SIG on Modeling Ensembles and Alternative Conformations**

[The third meeting of the CASP SIG on Modeling Ensembles and Alternative Conformations will be held on Wednesday June 7, 2023 at 11:00 am EDT \(NY time\) These meetings are held by zoom on first Wedne ...](#)

# 30 years!



# Võistlused võistlused - “tasuta” tööjõud?

Netflix challenge (\$1M prize for predicting star ratings for DVD-s)

[mlcontests.com/](http://mlcontests.com/) - ML võistlused (konverentsid jne)

[kaggle.com/](http://kaggle.com/) - **Kaggle**

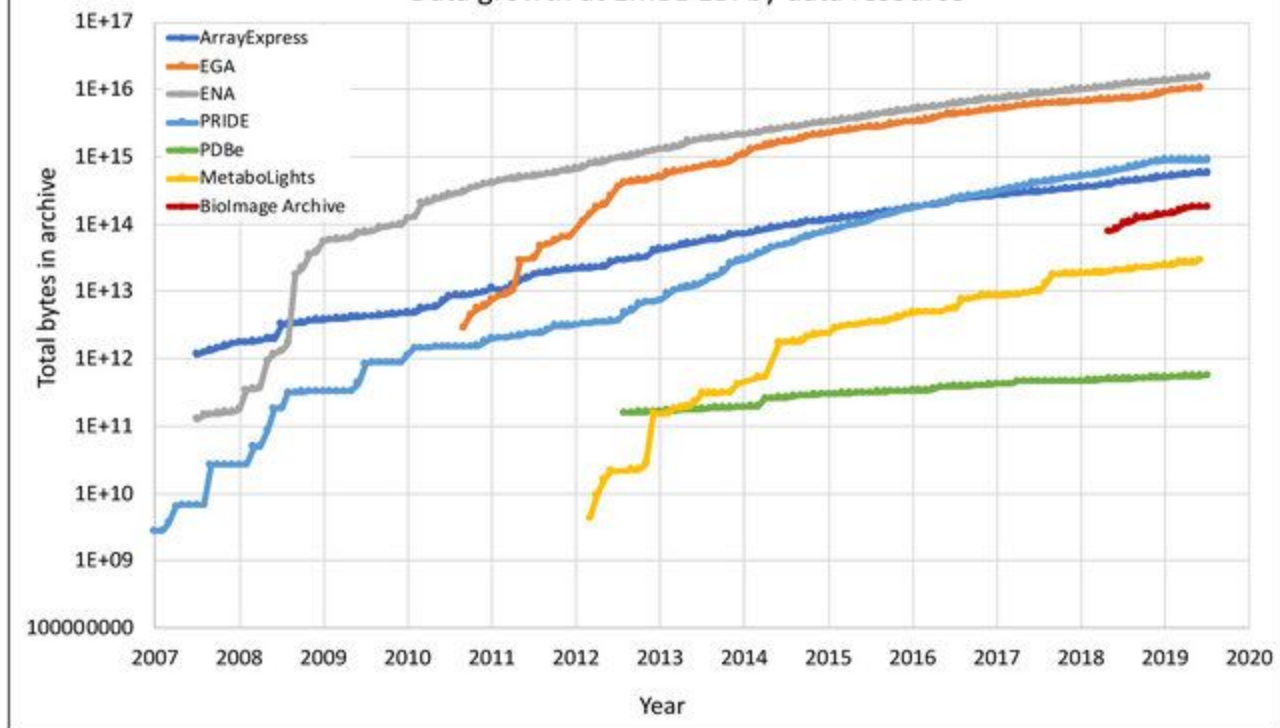
[machinehack.com/hackathons](http://machinehack.com/hackathons) - Machine Hack

[datahack.analyticsvidhya.com/contest/all/](http://datahack.analyticsvidhya.com/contest/all/) - Data Hack

[topcoder.com/challenges](http://topcoder.com/challenges) - **Top Coder**

<https://www.knowledgehut.com/blog/data-science/data-science-competitions>

Data growth at EMBL-EBI by data resource



[Published: 01 December 2001](#)

# Minimum information about a microarray experiment (MIAME)—toward standards for microarray data

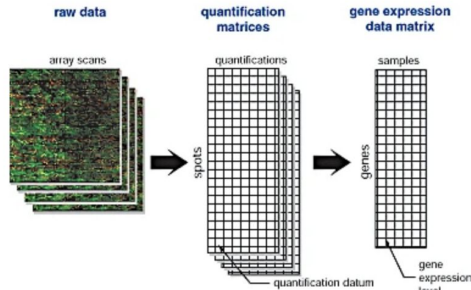
[Alvis Brazma](#) , [Pascal Hingamp](#), [John Quackenbush](#) , [Gavin Sherlock](#), [Paul Spellman](#), [Chris Stoeckert](#), [John Aach](#), [Wilhelm Ansorge](#), [Catherine A. Ball](#), [Helen C. Causton](#), [Terry Gaasterland](#), [Patrick Glenisson](#), [Frank C.P. Holstege](#), [Irene F. Kim](#), [Victor Markowitz](#), [John C. Matese](#), [Helen Parkinson](#), [Alan Robinson](#), [Ugis Sarkans](#), [Steffen Schulze-Kremer](#), [Jason Stewart](#), [Ronald Taylor](#), [Jaak Vilo](#) & [Martin Vingron](#)

[Nature Genetics](#) **29**, 365–371 (2001) | [Cite this article](#)

**20k** Accesses | **3001** Citations | **59** Altmetric | [Metrics](#)

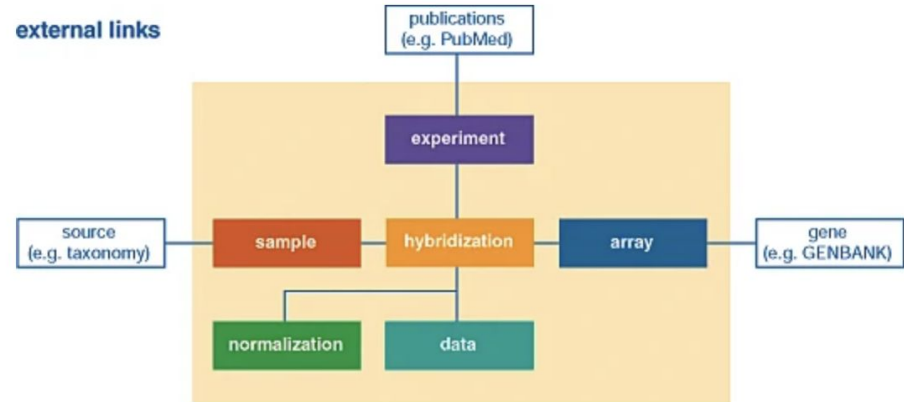
# Minimum information about a microarray experiment (MIAME)—toward standards for microarray data (2001)

Figure 2: Three levels of microarray gene expression data processing.



The raw data from microarray experiments are images. These images have to be quantified by image analysis software, which identifies spots related to each element on the array and measures the fluorescence intensity of each spot in each channel, together with the background intensity and a number of other quantifications, depending on the particular software (microarray quantification matrices). To obtain the final gene expression matrix, all the quantities related to each gene (either on the same array or on replicate arrays) have to be combined and the entire matrix has to be normalized to make different arrays comparable.

Figure 3

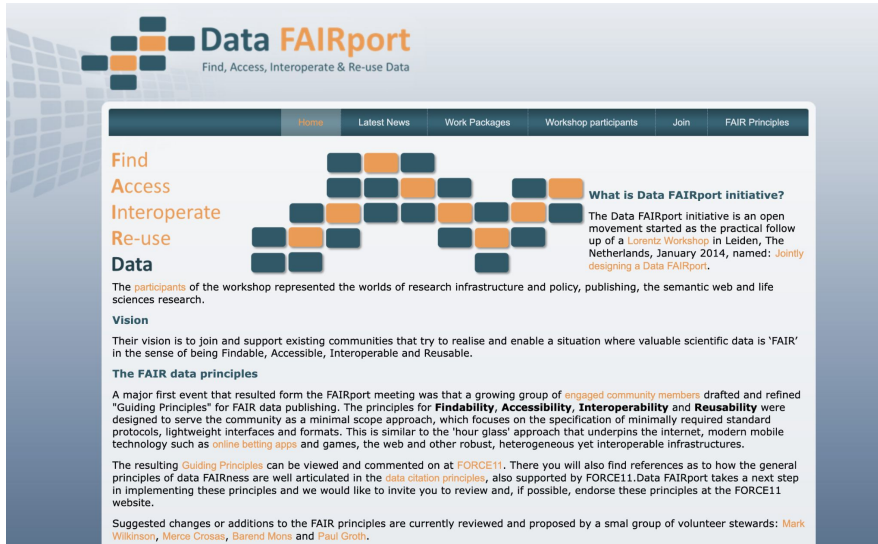


A schematic representation of six components of a microarray experiment.

**FAIR**

**F**indable **A**ccessible **I**nteroperable **R**eusable

# 2014 - Data Fairport



**Data FAIRport**  
Find, Access, Interoperate & Re-use Data

Home Latest News Work Packages Workshop participants Join FAIR Principles

**Find**  
**Access**  
**Interoperate**  
**Re-use**  
**Data**

**What is Data FAIRport initiative?**  
The Data FAIRport initiative is an open movement started as the practical follow up of a [Lorentz Workshop](#) in Leiden, The Netherlands, January 2014, named: *Jointly designing a Data FAIRport*.

The participants of the workshop represented the worlds of research infrastructure and policy, publishing, the semantic web and life sciences research.

**Vision**  
Their vision is to join and support existing communities that try to realise and enable a situation where valuable scientific data is 'FAIR' in the sense of being Findable, Accessible, Interoperable and Reusable.

**The FAIR data principles**  
A major first event that resulted from the FAIRport meeting was that a growing group of *engaged community members* drafted and refined "Guiding Principles" for FAIR data publishing. The principles for **Findability, Accessibility, Interoperability and Reusability** were designed to serve the community as a minimal scope approach, which focuses on the specification of minimally required standard protocols, lightweight interfaces and formats. This is similar to the 'hour glass' approach that underpins the internet, modern mobile technology such as *online betting apps* and games, the web and other robust, heterogeneous yet interoperable infrastructures.

The resulting *Guiding Principles* can be viewed and commented on at [FORCE11](#). There you will also find references as to how the general principles of data FAIRness are well articulated in the *data citation principles*, also supported by FORCE11. Data FAIRport takes a next step in implementing these principles and we would like to invite you to review and, if possible, endorse these principles at the FORCE11 website.

Suggested changes or additions to the FAIR principles are currently reviewed and proposed by a small group of volunteer stewards: [Mark Wilkinson](#), [Merce Crosas](#), [Barend Mons](#) and [Paul Groth](#).



Eesti on ELIXIR Euroopa  
asutajaliige;  
Rahvusvaheline teadusandmete taristu



**Data FAIRport**  
Find, Access, Interoperate & Re-use Data

Home Latest News Work Packages Works

## Participants

The following people participated in the Data Fairport inaugural meeting

Dr. Bengt Persson	ELIXIR Sweden, BILS
Dr. Erik van Mulligen	Erasmus Medical Centre, S&T
Dr. George Strawn	NCO/NITRD (USA)
Dr. Usbrand Jan Aalbersberg	Elsevier
Dr. Jan Willem Boiten	Dutch Techcentre for Life Sciences (DTL), CTMM
Dr. Mark Wilkinson	Technical University of Madrid, SADI
Dr. Maurice Bouwhuis	SURFSara
Dr. Michel Dumontier	Stanford, NCSO, Bio2RDF
Dr. Morris Swertz	Dutch Techcentre for Life Sciences (DTL), BBMRI-NL
Dr. Myles Axton	Nature Genetics
Dr. Niklas Blomberg	ELIXIR Hub (Europe)
Dr. Paul Groth	VU University of Amsterdam, W3C, Open PHACTS
Dr. Rene van Schaik	Netherlands eScience Center
Dr. Rob Hooft	Netherlands eScience Center, DTL
Dr. Ruben Kok	Dutch Techcentre for Life Sciences (DTL)
Dr. Scott Lusher	Netherlands eScience Center
Dr. Ted Slater	YarcDATA
Dr. Thierry Sengstag	Swiss Institute of Bioinformatics (SIB)
Drs. Albert Mons	Phortos Consultants, Euretos
Drs. Jan Velterop	Acknowledge
Drs. Mr. Arie Baak	Phortos Consultants, Euretos
Gaby Appleton	Elsevier
Olivier Dumon	Elsevier
Prof. Abel L. Packer	SciELO
Prof. Anthony Brookes	University of Leicester, GEN2PHEN-Alliance & BioShARe
Prof. Barend Mons	Leiden University Medical Center, ELIXIR NL
Prof. Carole Goble	University of Manchester, UK ELIXIR node.
Prof. Frank van Harmelen	VU University Amsterdam, Open PHACTS
Prof. Jaap Heringa	VU University of Amsterdam, ELIXIR NL
Prof. Johan van der Lei	Erasmus Medical Center, EMIF
Prof. Joost Kok	Leiden University, LIACS
Prof. Tim Clark	Harvard medical school, Mass. General Hospital, Force11



**F**indable

Data and materials enriched with metadata assigned with a unique identifier



**A**ccessible

Data and metadata stored in a trusted repository with an open and free protocol. Accessible by machines and humans



**I**nteroperable

Using vocabularies and public domain ontologies the metadata can be referenced and linked



**R**eusable

Additional documentation and protocols describing the acquisition of the data, licensed with a detailed provenance



## Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

**F1. (Meta)data are assigned a globally unique and persistent identifier.**

**F2. Data are described with rich metadata (defined by R1 below).**

**F3. Metadata clearly and explicitly include the identifier of the data they describe.**

**F4. (Meta)data are registered or indexed in a searchable resource.**

## Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

**A1. (Meta)data are retrievable by their identifier using a standardised communications protocol.**

**A1.1 The protocol is open, free, and universally implementable.**

**A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.**

**A2. Metadata are accessible, even when the data are no longer available.**

## Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.**
- I2. (Meta)data use vocabularies that follow FAIR principles.**
- I3. (Meta)data include qualified references to other (meta)data.**

## Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.







**R1. (Meta)data are richly described with a plurality of accurate and relevant attributes.**


**R1.1. (Meta)data are released with a clear and accessible data usage license.**

**R1.2. (Meta)data are associated with detailed provenance.**

**R1.3. (Meta)data meet domain-relevant community standards.**

avaandmed.eesti.ee

 Andmestikud  Teabevaldajad  Kasutuslood  Statistika  Juhendid  Andmesoovid

Eesti avaandmed  Arutelud Logi sisse Loo konto et ▾

# Eesti avaandmete teabevärv

Avaandmed on tasuta andmed, mis on kõigile antud vabalt kasutamiseks, taaskasutamiseks ja jagamiseks ning mida saab kasutada äriliste ja mittetulunduslike ettevõtmiste käivitamiseks, uuringute läbiviimiseks ja andmepõhiste otsuste tegemiseks.

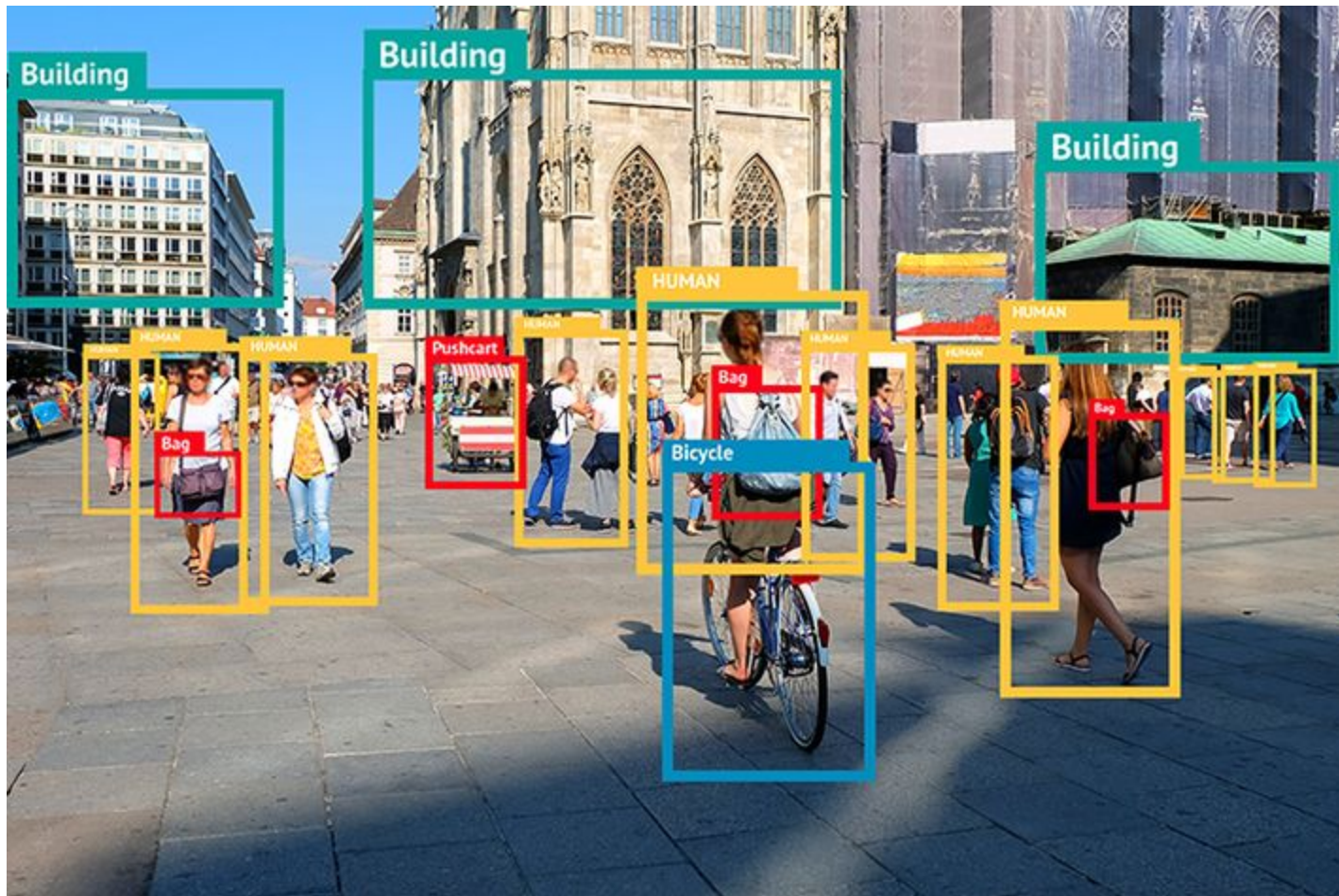
Avaandmete teabevärv pakub kõigile võimaluse avaandmeid tarbida ning visualiseerida, koondades endas ka avaandmete põhjal loodud kasutuslood.

[Vaata andmestikke](#)

1779 andmestikku 2230 teabevaldajat <sup>1</sup>

 <b>Tervis</b> 	 <b>Keskond</b> 	 <b>Põllumajandus, kalandus, metsandus ja toiduained</b> 	 <b>Majandus ja rahandus</b> 
 <b>Õigusemõistmine, õigussüsteem ja avalik turvalisus</b> 	 <b>Valitsus ja avalik sektor</b> 	 <b>Elanikkond ja ühiskond</b> 	 <b>Transport</b> 
 <b>Haridus, kultuur ja sport</b> 	 <b>Energeetika</b> 	 <b>Piirkonnad ja linnad</b> 	 <b>Teadus ja tehnoloogia</b> 

Pildid



# Research in production (PerkinElmer Harmony)

The screenshot displays the Harmony software interface for image analysis. The main window shows a microscopy image with numerous nuclei highlighted in various colors (red, green, blue, yellow, orange, purple). The interface is divided into several panels:

- Top Panel:** Navigation icons for Setup, Run Experiment, Image Analysis, Evaluation, Settings, and Help.
- Left Panel:** Analysis configuration for "Nuclei Prediction".
  - Analysis: \*\_Sequence\_For\_Input\_\*
  - Measurement: Seven\_Cell\_Lines\_Hoechst...
  - Analysis Sequence: Input Image, Using: Individual Planes, FFC None
  - ABB: Find Nuclei DNN
  - Status: Under Development
  - Channel: Brightfield
  - Model: UNet
  - Threshold: 0.5
  - Minimum Size: 10  $\mu\text{m}$
  - Upper Threshold: 0
  - Output Population: Nuclei
  - Output Image: Nuclei Prediction
  - Define Results: Output: 1 Well Results, 0 Object Results
- Image Control Panel:** Controls for the image, including Coloring (Highlight), Show Scale, Channels (Brightfield), Color (Gray), Auto Contrast (1.50), HOECHST 33342, and Nuclei Prediction.
- Navigation Panel:** Plate and Well selection. The plate is labeled "Seven\_Cell\_Lines\_Ho..." and the well is highlighted.
- Image Analysis Results Panel:** Summary and Properties for the Nuclei population.

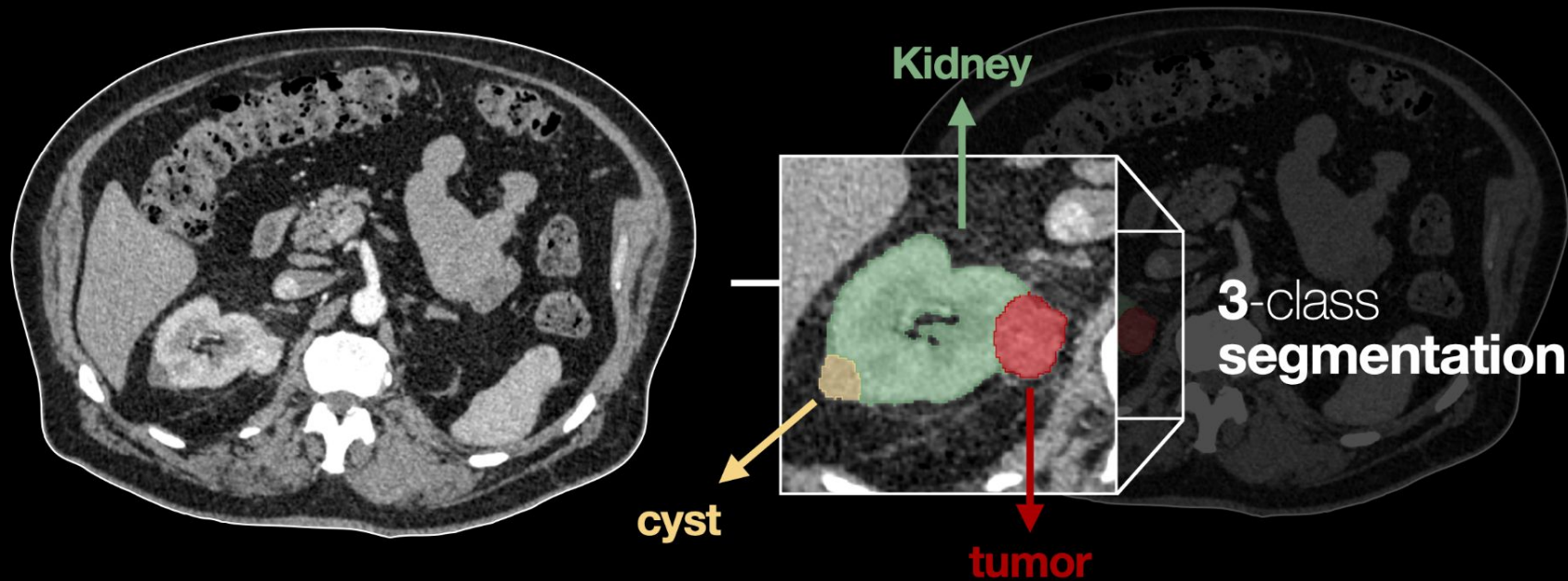
Population: Nuclei		Value						
Number of Objects	146							
Property	Mean	CV %	StdDev	Median	Max	Min	Sum	
Size [ $\mu\text{m}$ ]	14.8599	17.8496	2.65228	14.7232	21.6502	10.0607	2169.54	



# The task

(based on data from Tartu Ülikooli Kliinikum)

We are testing a model that can detect **kidneys** and kidney **lesions** and **cysts**





## Coronavirus analysis tools developed by scientists at the University of Tartu:

**TARTU ÜLIKOOL**  
 Demographic, geographic and epidemiological study of COVID-19 susceptible citizens

Introduction Return Main information My data

Entry: 2020-03-31 13:07

Basic information Place of residence Symptoms Onset or change of your symptoms Encounters Health data

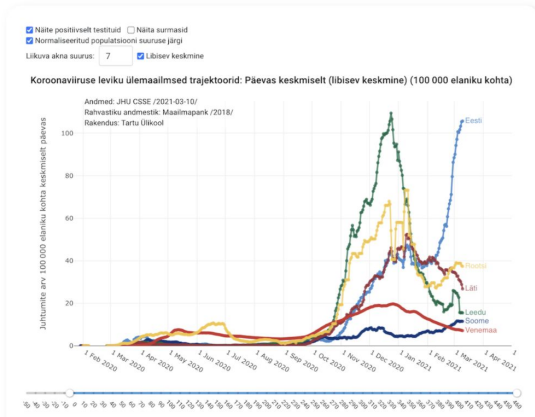
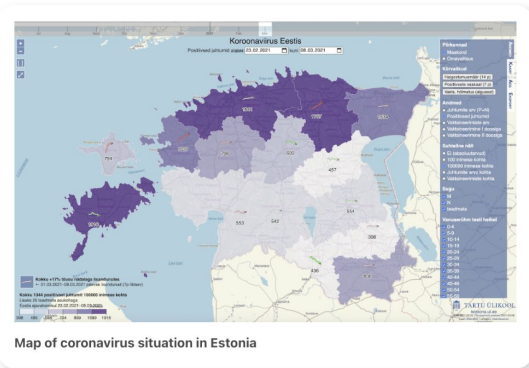
**PARTICIPATE IN THE SURVEY AND FILL OUT THE HEALTH DIARY HERE**

The highest number of participants in the last 24 hours, if you have an opportunity to participate in the study.

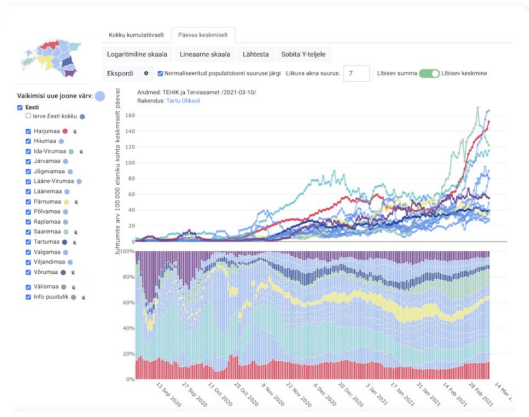
Symptoms

- No symptoms
- Cough
- Sputum secretion
- Headache
- Sore throat
- Stomach ache

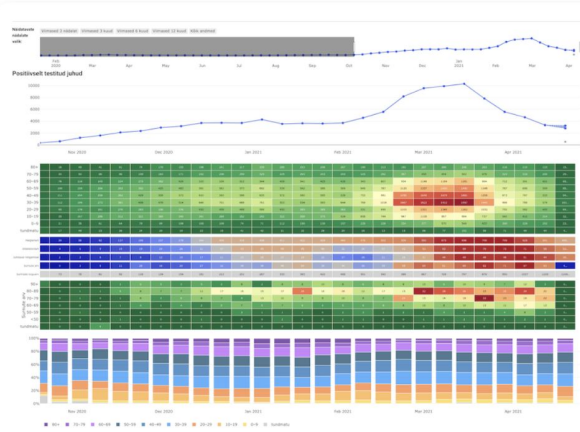
Coronavirus dissemination study



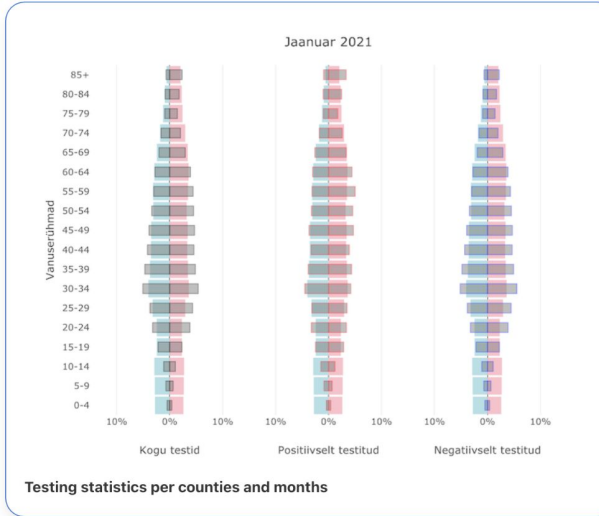
Worldwide comparison of COVID-19 infection trajectories [web application]



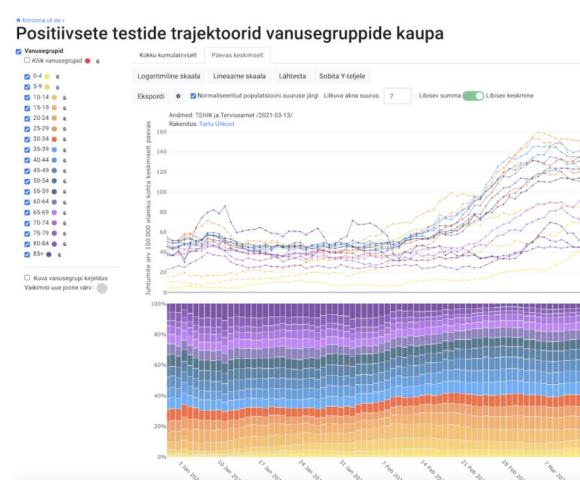
Comparison of COVID-19 trajectories in Estonia



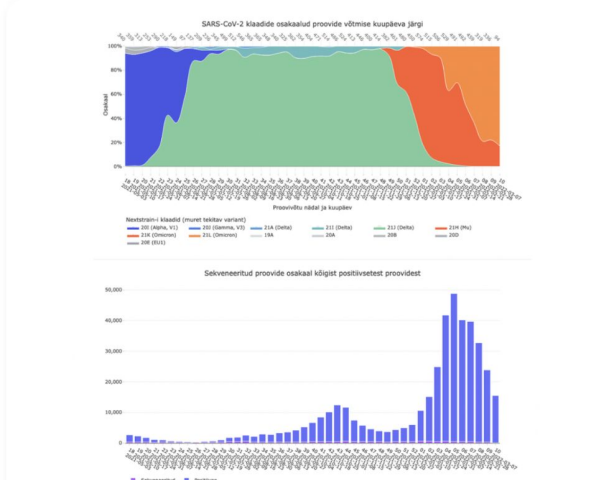
Weekly PCR-testing, hospitalisation and death statistics per age groups



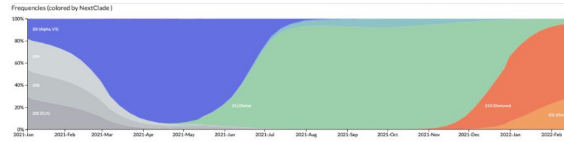
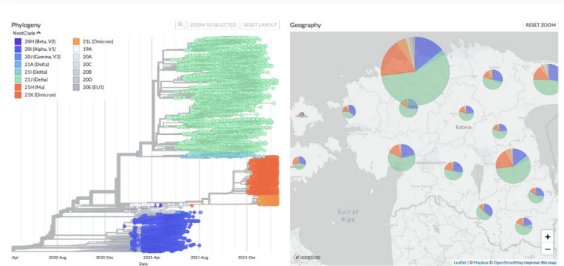
Testing statistics per counties and months



Comparison of COVID-19 trajectories by age groups in Estonia

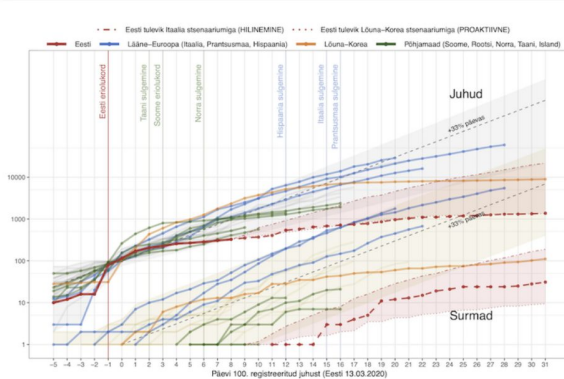


SARS-CoV-2 lineages in Estonia

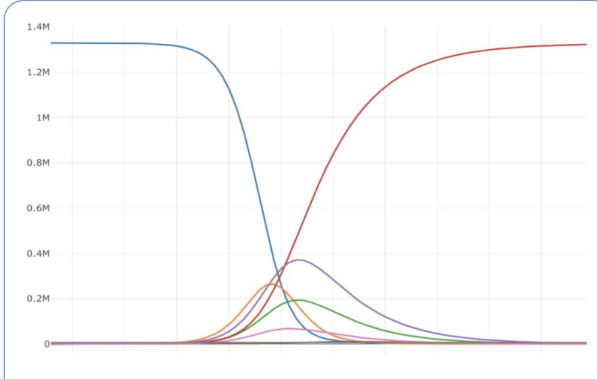


Auspice visualisations of SARS-CoV-2 evolution

Estonian COVID-19 Data Portal



Coronavirus infection trajectories on 26.03.2020, 16.04.2020 [PDF]



SEIAHR prognostic model for predicting spread of the infection

This work has been supported by the kind donation of Väino Kaldoja

# Media coverage



## Postimees

April 3, 2020

Otse Postimehest: Eesti on koroonaviiruse uuringutes maailmas esirinnas



## Press release

April 3, 2020

University of Tartu researchers study the spread of coronavirus in Estonia



## ETV

April 3, 2020

Koroonauuring



## ETV

April 5, 2020

Aktuaalne kaamera. 2020 Nädal (22:20)



## Postimees

April 7, 2020

Tartu Ülikooli professorid: Covid-19 lõppu on raske prognoosida



## Stolitsa

April 20, 2020

Jaak Vilo, Hedi Peterson: SARS-CoV-2 spreads more easily than flu



## Life in Estonia Magazine

June 26, 2020

Study on Covid-19: Predicting possible scenarios (pp. 40–42)



## Tartu Postimees

June 26, 2020

Tartu Ülikooli arvutiteadlased arendavad rahvusvahelist koostööd koroonaviiruse leviku uurimiseks



## Press release

June 26, 2020

Tartu Ülikooli arvutiteadlased arendavad rahvusvahelist koostööd koroonaviiruse leviku uurimiseks



## Tervis

October 16, 2020

Tartu Ülikool jätkab koroonaviiruse leviku uuringut eraannetuse toel



## Forte

October 16, 2020

Tuntud ettevõtja annetas koroonauuringuteks suure summa – ülikool tänab



## Tartu Postimees

October 16, 2020

Tartu ülikool sai suure eraannetuse koroonaviiruse leviku andmete uurimiseks



## Press release

October 16, 2020

Tartu Ülikool jätkab koroonaviiruse leviku andmete analüüsi suure eraannetuse toel



## Tartu Postimees

October 30, 2020

Aare Abro: koroonat saab veel peatada, aga varsti võib olla juba liiga hilja



## EPL Delfi

November 21, 2021

INFOLAENG | Millises Eesti piirkonnas tuleb kõige koroonavabam talv?

# Ülevaade eriarstiabi saanute arvust ja kuludest vanuse ning põhidiagnooside lõikes

2019 ▾

€

👤

👤

👤

👤👤

Per capita

Kokku

731M

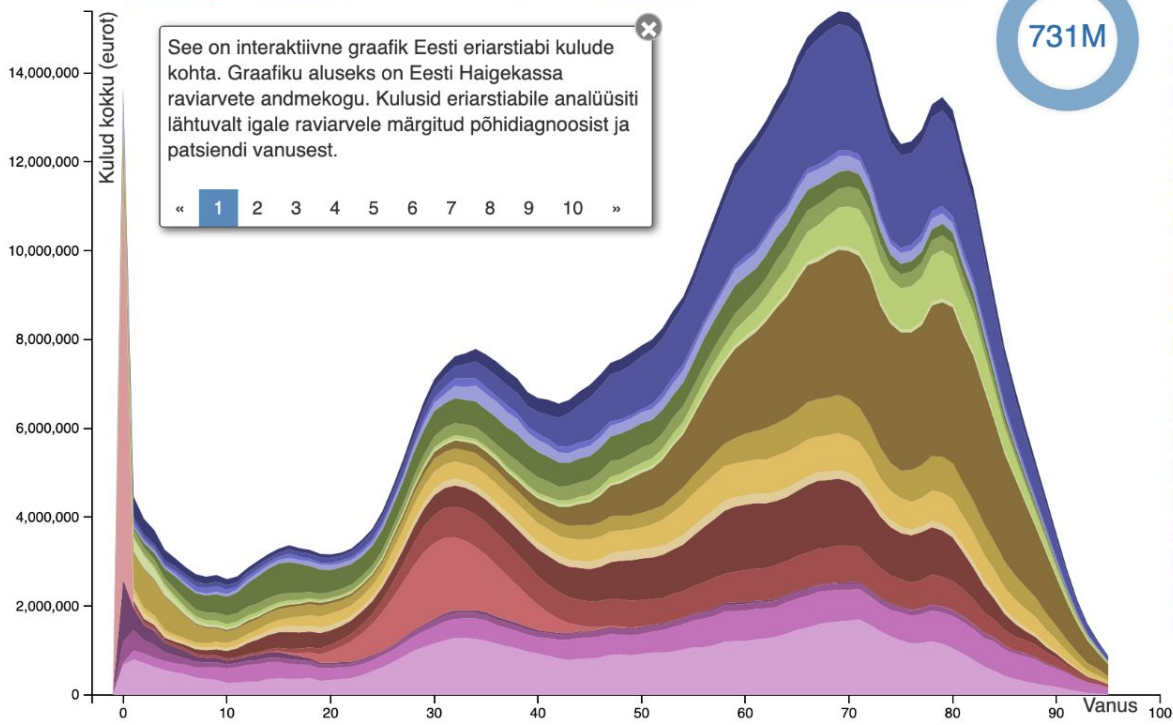
Info

Et

En

See on interaktiivne graafik Eesti eriarstiabi kulude kohta. Graafiku aluseks on Eesti Haigekassa raviarvete andmekogu. Kulud eriarstiabile analüüsiti lähtuvalt igale raviarvele märgitud põhidiagnoosist ja patsiendi vanusest.

« 1 2 3 4 5 6 7 8 9 10 »



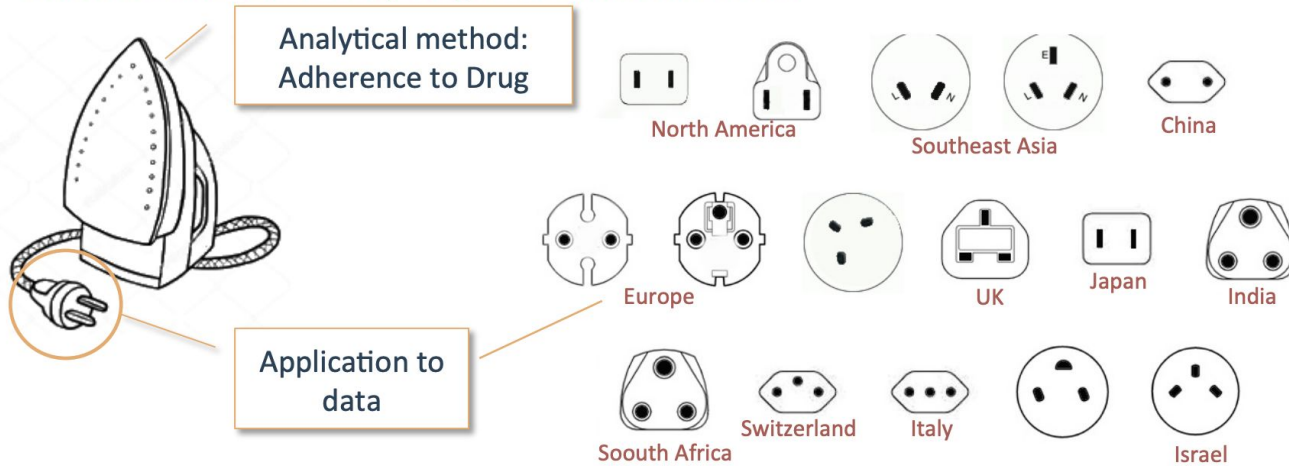
- Teatavad nakkus- ja parasiithaigused
- Kasvajad
- Vere- ja vereloomeelundite haigused ning teatavad immuunmehhanismidega seotud haigusseisundid
- Sisesekreetsiooni-, toitumis- ja ainevahetushaigused
- Psüühika- ja käitumishäired
- Närvisüsteemahaigused
- Silma- ja silmamanuste haigused
- Kõrva- ja nibujätkehaigused
- Vereringeelundite haigused
- Hingamiselundite haigused
- Seedeelundite haigused
- Naha- ja nahaaluskoe haigused
- Lihaskonna ja sidekoehaigused
- Kuse-suguelundite haigused
- Rasedus, sünnitus ja sünnitusjärgne periood
- Perinataal- e sünniperioodis tekkinud teatavad seisundid
- Kaasasündinud vääramendid, deformatsioonid ja kromosoomianomaaliad
- Mujal klassifitseerimata sümptomid, tunnused ja kliiniliste ning laboratoorse teadmiste hälbed
- Vigastused, mürgistused ja teatavad muud välispõhjuste toime tagajärjed
- Tervise seisundit mõjustavad tegurid ja kontaktid tervise teenistusega

Allikas: Eesti Haigekassa raviarvete andmekogu 2013-2019

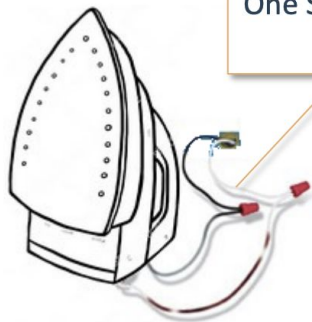


# Current Approach: "One Study – One Script"

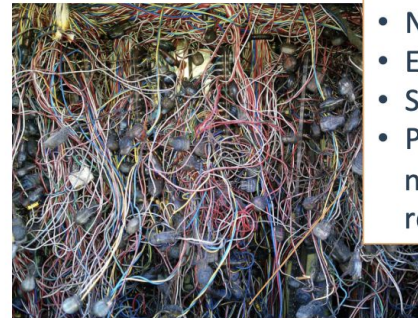
"What's the adherence to my drug in the data assets I own?"



Current solution:



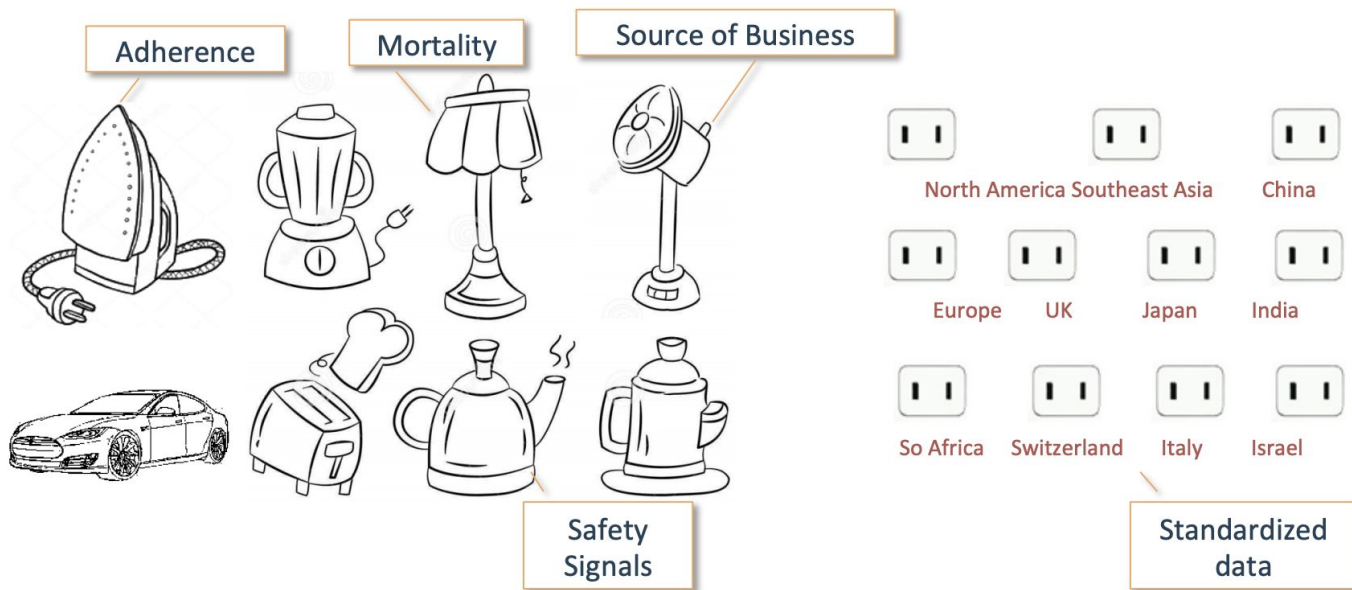
One SAS or R script  
for each study



- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use



# Solution: Data Standardization Enables Systematic Research

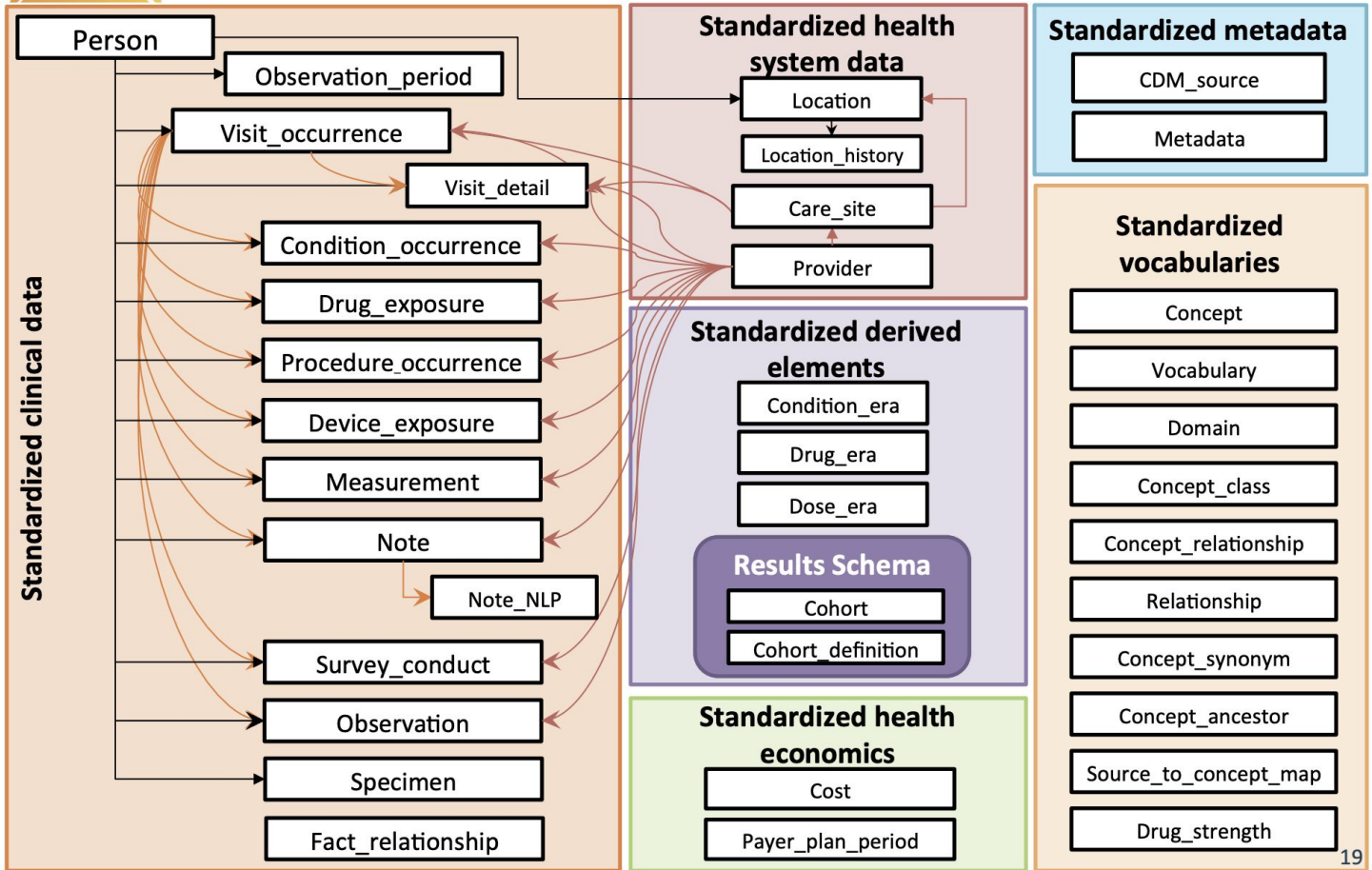


OHDSI Tools

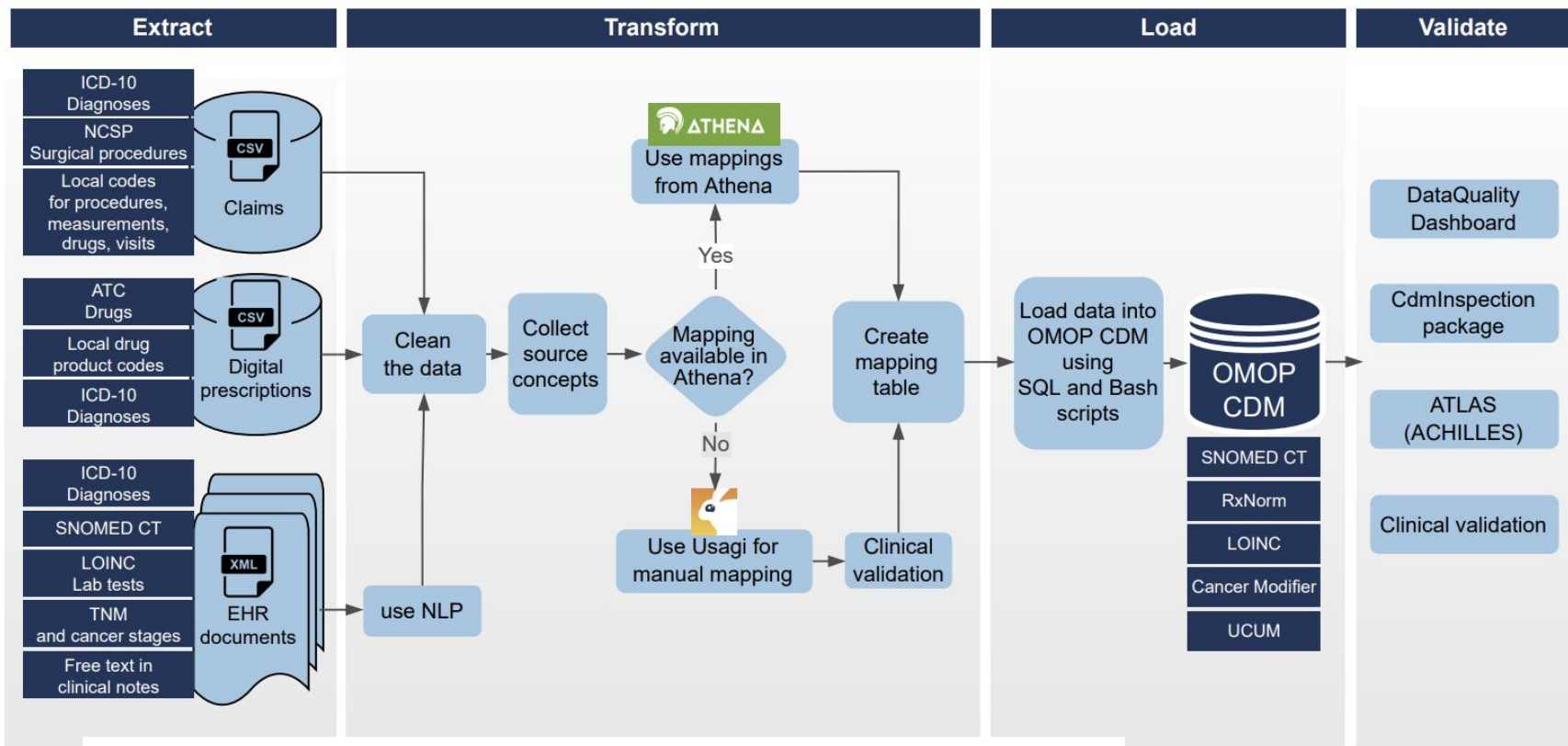
OMOP CDM



# CDM Version 6 Key Domains







## Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned

Marek Oja, Sirli Tamm, Kerli Mooses, Maarja Pajusalu, Harry-Anton Talvik, Anne Ott, Marianna Laht, Maria Malk, Marcus Lõo, Johannes Holm, Markus Haug, Hendrik Šuvalov, Dage Särg, Jaak Vilo, Sven Laur, Raivo Kolde, Sulev Reisberg

doi: <https://doi.org/10.1101/2023.02.16.23285697>

medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

# RESEARCH GROUP OF HEALTH INFORMATICS



We believe that **data-informed** healthcare more efficient, patient-centered, and we work passionately to **generate** **epidemiological studies**, doing **software** in the healthcare domain.

We have **long-term experience** **in the region** for health data analysis and processing.

Furthermore, we like to **share our knowledge** with the academic community and the public. Interact with our bachelor's and master's students.

teatmik.ee

statistika.ut.ee

haridussilm.ee

etis.ee/statistika

andmed.stat.ee

Anna Aljanaki "Software of my Head and Hardware of Delta" 2020





# Kokkuvõte - andmed on tähtsad **kuid neid on palju**

**Andmed on otsustamisel tähtsad**

**Avatud andmed võimaldavad :**

taaskasutust ja läbipaistvust otsustamisel

Teaduse andmed on sageli väga kallid, neid tuleb analüüsida mitmest nurgast

**Üleskutse üliõpilaste meelitamiseks**

- Kui on soov kaasata, tuleks poolele teele vastu tulla  
n. valmistada ette andmestikke ja küsimusi  
käesoleva seminari loengute salvestused
- Stipid, praktikakohad, tööstusmagistrantuur
- Võistlused ja üleskutsed - rahalised preemiad jne.



DELTA  
TARTU ÜLIKOOL

liikumisvõime  
sõidukiha teater  
TARTU ÜLIKOOL

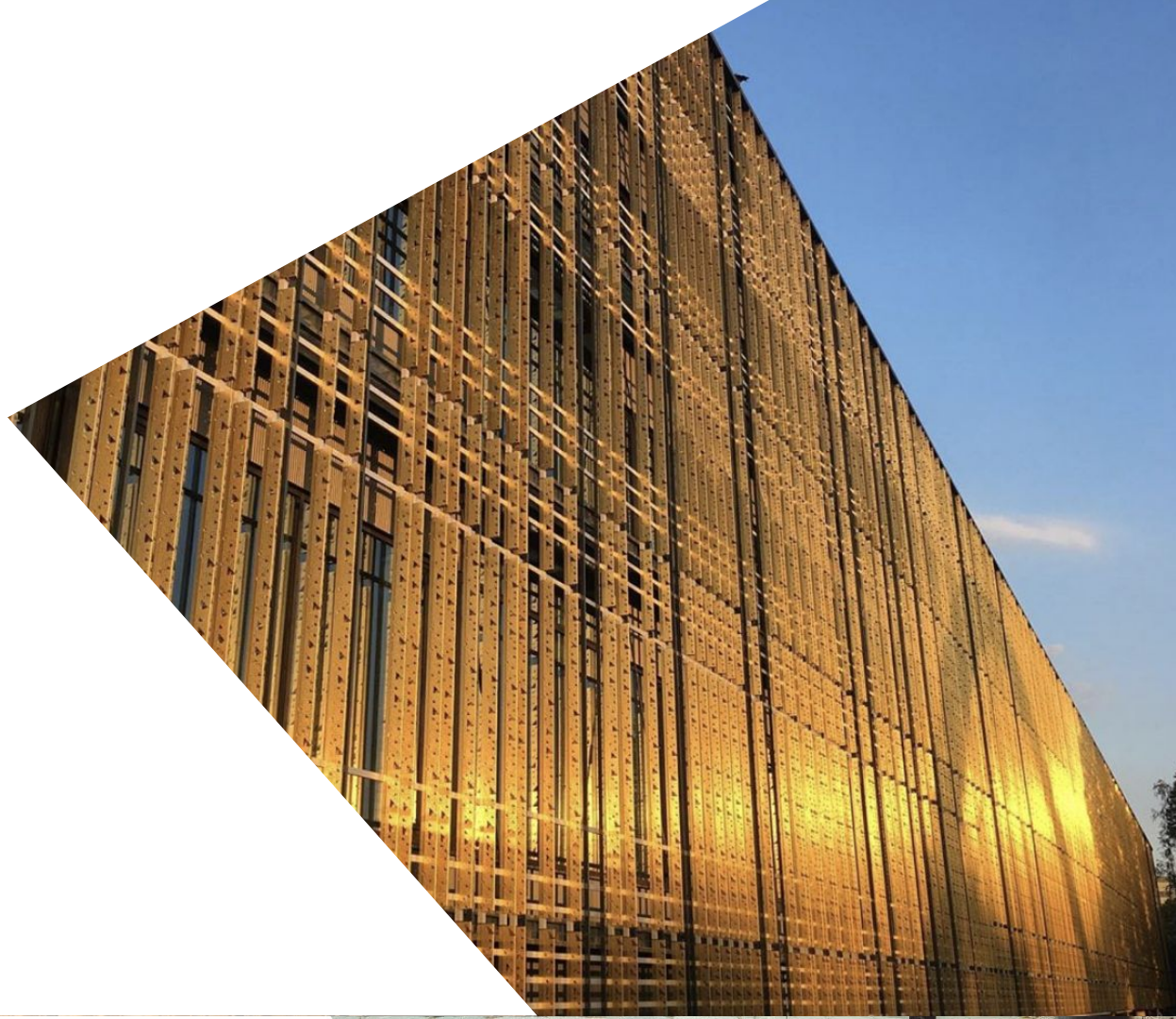
**Bolt**

**TÜ arvutiteaduse  
instituut**

**Aitäh  
kuulamast!**

**[vilo@ut.ee](mailto:vilo@ut.ee)**

**[cs.ut.ee](http://cs.ut.ee)**







# All students: IT has nearly 2x more than natural sciences

## Üliõpilaste arv rahastusallika ja õppevaldkonna lõikes

Õppevaldkond	15/16	16/17	17/18	18/19	19/20	20/21	21/22	22/23
⊕ Haridus	3,523	3,272	3,137	3,162	3,257	3,422	3,532	3,661
⊕ Humanitaaria ja kunstid	6,468	6,082	6,096	6,167	6,068	6,075	6,071	6,034
⊕ Informatsiooni- ja kommunikatsioonitehnoloogiad	4,170	4,155	4,059	4,377	4,569	4,860	4,780	4,790
⊕ Loodusteadused, matemaatika ja statistika	3,169	2,929	2,825	2,698	2,619	2,726	2,812	2,805
⊕ Põllumajandus, metsandus, kalandus ja veterinaaria	1,150	1,050	971	947	942	1,023	1,065	1,096
⊕ Sotsiaalteadused, ajakirjandus ja teave	3,820	3,438	3,011	2,848	2,867	2,828	2,849	2,853
⊕ Teenindus	2,967	2,746	2,682	2,529	2,449	2,435	2,455	2,414
⊕ Tehnika, tootmine ja ehitus	8,590	7,793	7,293	6,877	6,661	6,587	6,265	6,041
⊕ Tervis ja heaolu	5,563	5,492	5,428	5,691	5,873	5,944	6,084	6,104
⊕ Ärindus, haldus ja õigus	11,672	10,836	10,652	10,519	9,873	9,359	8,698	8,333
<b>Total</b>	<b>51,092</b>	<b>47,793</b>	<b>46,154</b>	<b>45,815</b>	<b>45,178</b>	<b>45,259</b>	<b>44,611</b>	<b>44,131</b>

# PhD students: ICT has 2x less than natural sciences

## Üliõpilaste arv rahastusallika ja õppevaldkonna lõikes

Õppevaldkond	15/16	16/17	17/18	18/19	19/20	20/21	21/22	22/23
⊕ Haridus	114	101	98	99	99	95	96	88
⊕ Humanitaaria ja kunstid	419	387	417	431	401	394	386	408
⊕ Informatsiooni- ja kommunikatsioonitehnoloogiad	219	206	191	184	188	214	250	249
⊕ Loodusteadused, matemaatika ja statistika	690	648	609	577	543	532	526	514
⊕ Põllumajandus, metsandus, kalandus ja veterinaaria	135	129	125	128	134	139	129	128
⊕ Sotsiaalteadused, ajakirjandus ja teave	326	294	208	200	195	184	187	177
⊕ Teenindus	28	25	22	24	21	19	20	15
⊕ Tehnika, tootmine ja ehitus	419	372	355	331	316	332	350	321
⊕ Tervis ja heaolu	217	220	225	238	232	239	240	229
⊕ Ärindus, haldus ja õigus	266	252	240	200	187	169	169	163
<b>Total</b>	<b>2,833</b>	<b>2,634</b>	<b>2,490</b>	<b>2,412</b>	<b>2,316</b>	<b>2,317</b>	<b>2,353</b>	<b>2,292</b>

**Table 3.7. Funding of all active (i.e. ongoing) research projects according to the Estonian Research Information System in 2018–2021 and the ratio of funding in all research fields (%) according to the Frascati classification. The main research grants awarded by the Estonian Research Council (PUT, IUT, PRG, PSG grants) have been separately highlighted, as well as all other grant types, contractual and corporate projects, including foreign financed projects.**

	Research grants awarded by the Estonian Research Council (PUT, IUT, PRG, PSG grants) (million EUR)					All active research projects in Estonian Research Information System (incl. foreign, contractual and corporate financed) (million EUR)					Ratio (all active projects/research grants awarded by Estonian Research Council) 2021
	2018	2019	2020	2021	Share (2021)	2018	2019	2020	2021	Share (2021)	
<b>1. Natural Sciences</b>	<b>18.5</b>	<b>18.4</b>	<b>21.2</b>	<b>21.9</b>	<b>48.4%</b>	<b>38.2</b>	<b>48.6</b>	<b>50.3</b>	<b>50.6</b>	<b>45.4%</b>	<b>2.3</b>
1.1. Mathematics	0.6	0.5	1.1	1.2	2.6%	0.0	0.0	0.4	0.5	0.4%	0.4
1.2. Computer and information sciences	1.6	1.3	1.6	1.8	3.9%	7.4	12.0	14.2	12.7	11.3%	7.1
1.3. Physical sciences	3.4	2.9	4.1	4.0	8.9%	3.7	4.0	4.2	4.7	4.2%	1.2
1.4. Chemical sciences	2.4	3.1	3.4	2.9	6.4%	5.5	4.8	4.6	5.8	5.2%	2.0
1.5. Earth and related environmental sciences	2.3	2.3	2.5	1.9	4.1%	5.5	8.3	8.9	8.4	7.5%	4.5
1.6. Biological sciences	8.2	8.3	8.6	10.2	22.6%	16.1	19.4	17.9	18.6	16.7%	1.8
<b>2. Engineering and technology</b>	<b>4.7</b>	<b>4.9</b>	<b>6.2</b>	<b>6.4</b>	<b>14.1%</b>	<b>19.0</b>	<b>22.5</b>	<b>25.8</b>	<b>23.5</b>	<b>21.0%</b>	<b>3.7</b>
2.1. Civil Engineering	0.4	0.3	0.5	0.6	1.2%	2.9					
2.2. Electrical engineering, electronic engineering, information engineering	1.2	1.4	1.4	1.6	3.5%	7.8					
2.3. Mechanical engineering			0.2	0.2	0.5%	0.4					

ICT 12x less Estonian grant funding

University of Tartu

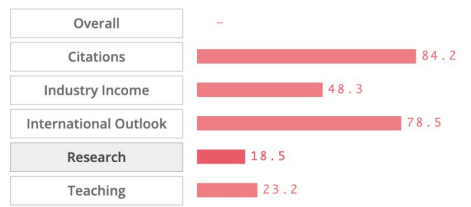
Official website

BREAKDOWN OF RANKING:

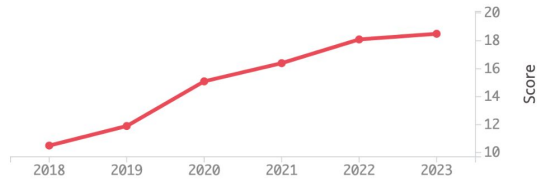
Computer Science

2023 Rank: 176-200

[View full ranking table](#)



Breakdown via year: RESEARCH



Ranking position 2018 to 2023:

2018	2019	2020	2021	2022	2023
201-250	251-300	201-250	176-200	201-250	176-200

SELECT FROM ONE OF THE BELOW RANKINGS:

<p>World University Rankings 2024 Rank: 301-350</p> <p>INDUSTRY 64.1</p>	<p>Arts &amp; humanities 2023 Rank: 251-300</p> <p>OVERALL -</p>	<p>Clinical, pre-clinical &amp; health 2023 Rank: 176-200</p> <p>OVERALL -</p>
<p>Life sciences 2023 Rank: 151-175</p> <p>OVERALL -</p>	<p>Physical sciences 2023 Rank: 251-300</p> <p>OVERALL -</p>	<p>Social sciences 2023 Rank: 301-400</p> <p>OVERALL -</p>
<p>Business &amp; Economics 2023 Rank: 201-250</p> <p>OVERALL -</p>	<p>Computer Science 2023 Rank: 176-200</p> <p>RESEARCH 18.5</p>	<p>Education 2023 Rank: 251-300</p> <p>OVERALL -</p>

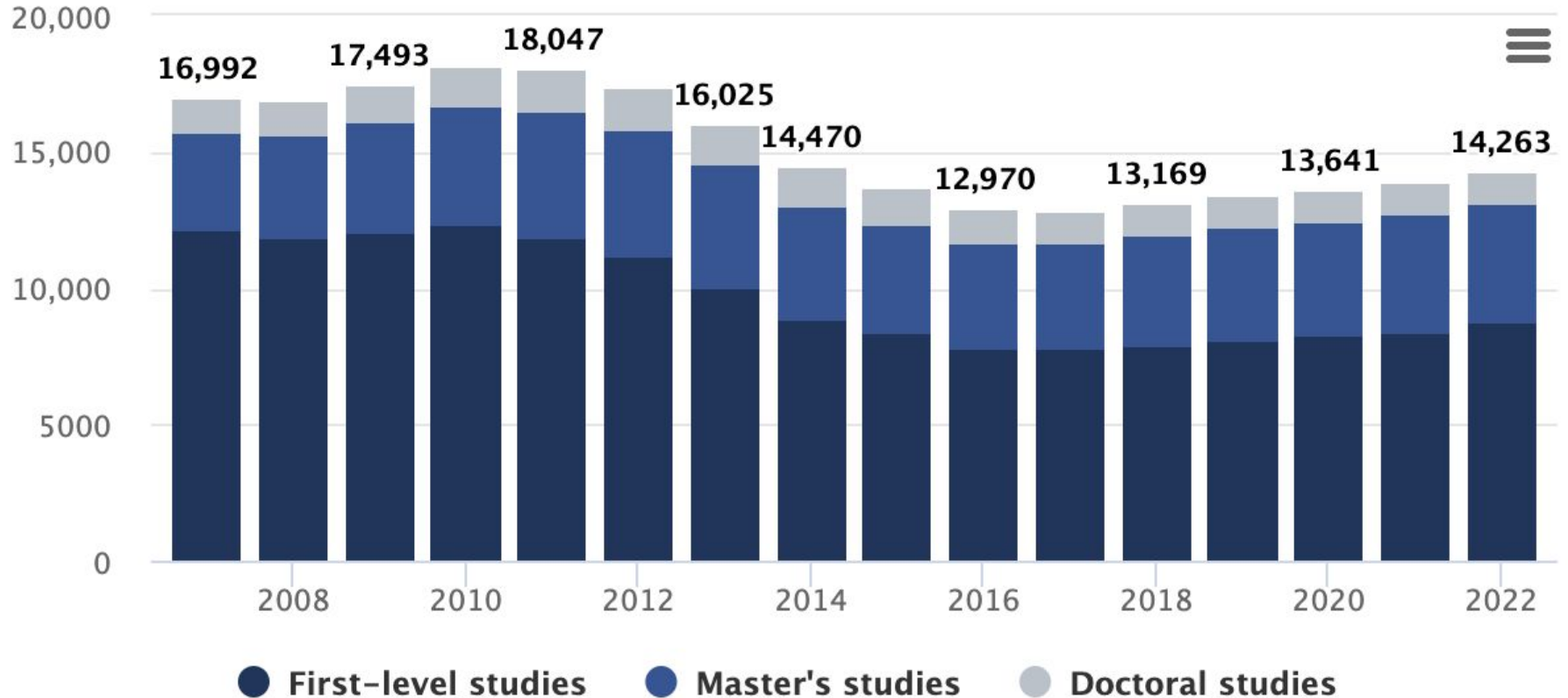
# UniTartuCS

## 176-200

# Tartu University students: 18K -> 14K (Estonia: 70K -> 44K)

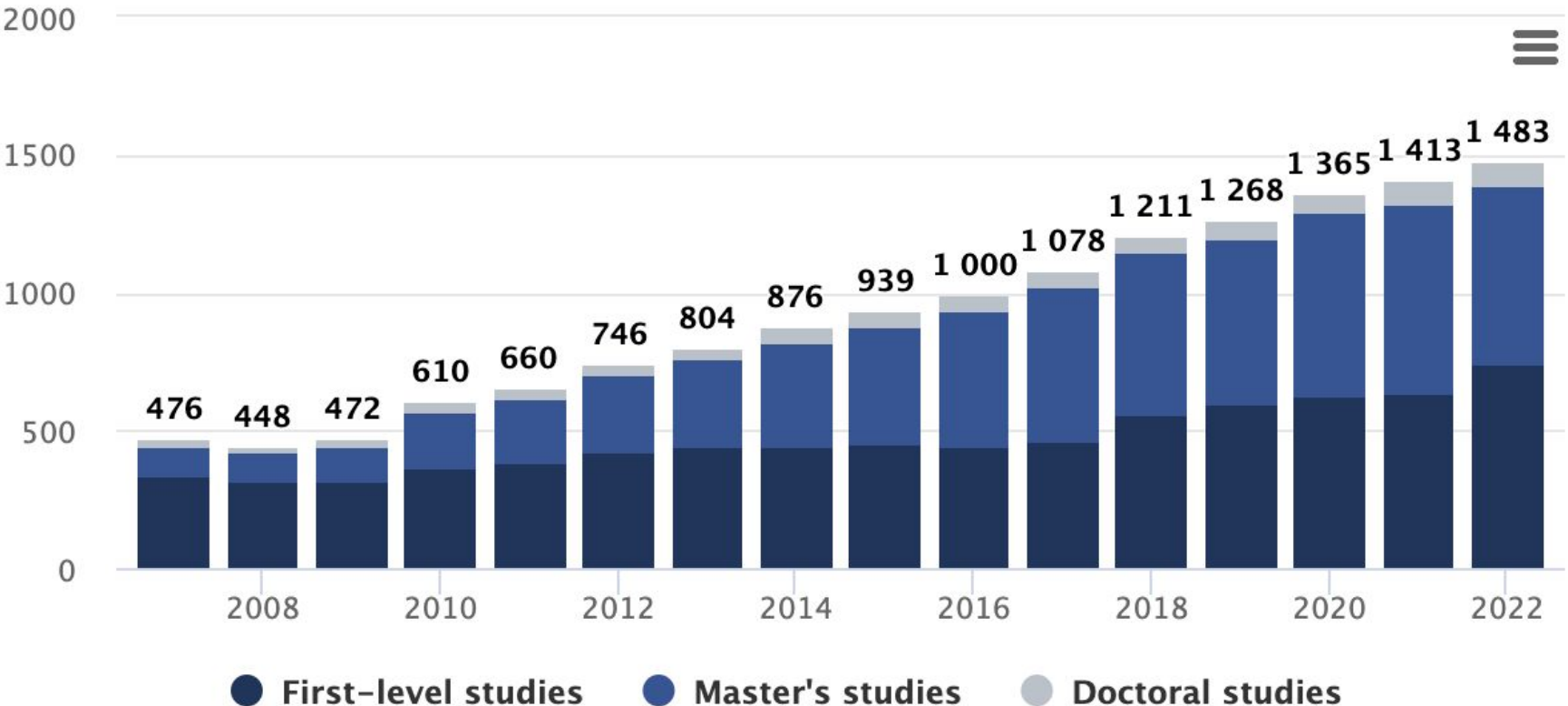
Estonian students     International students

statistika.ut.ee  
haridussilm.ee



# UniTartu CS students growth BSc, MSc, PhD

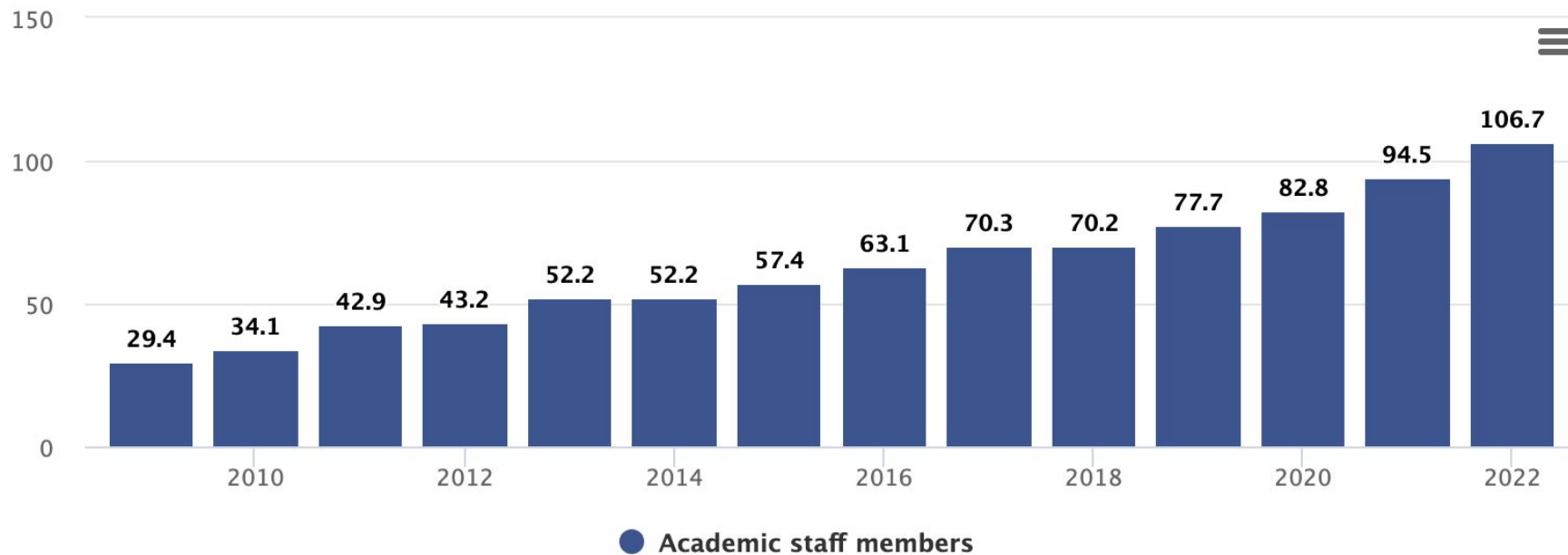
Estonian students  International students



## Number of employees (FTE)

Academic staff members  Support staff

International staff only

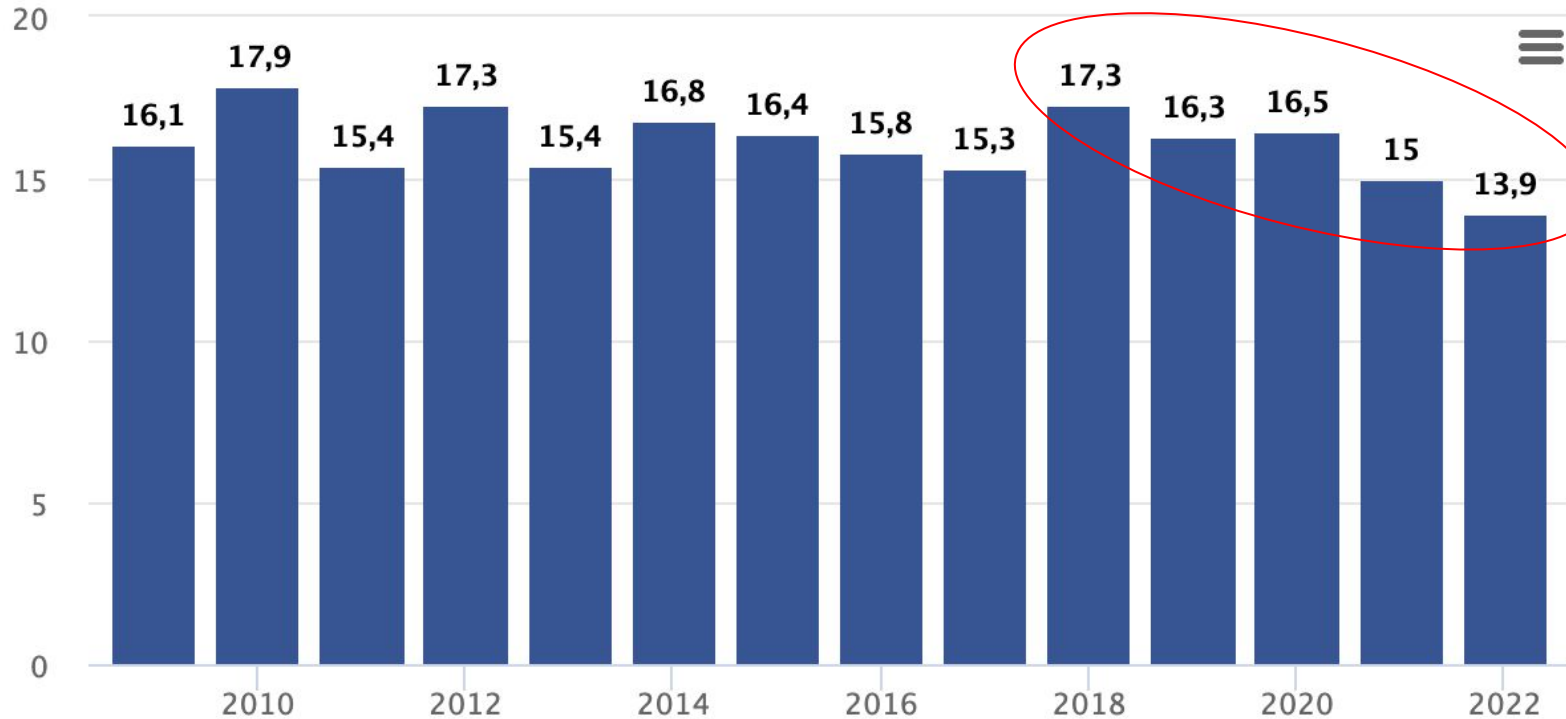




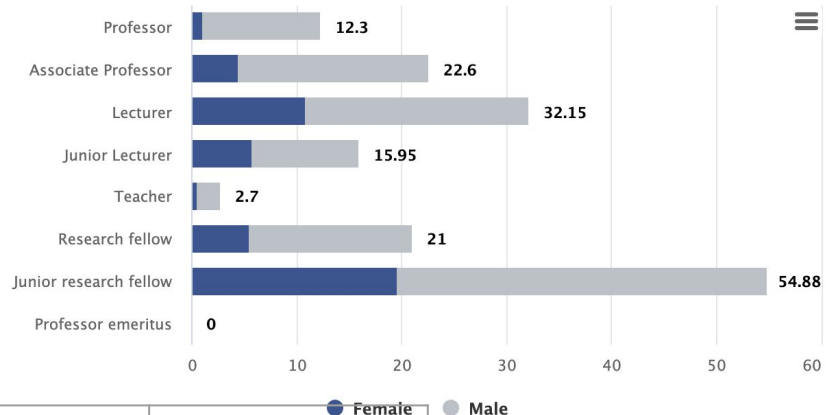
# Number of students per academic staff member

statistika.ut.ee

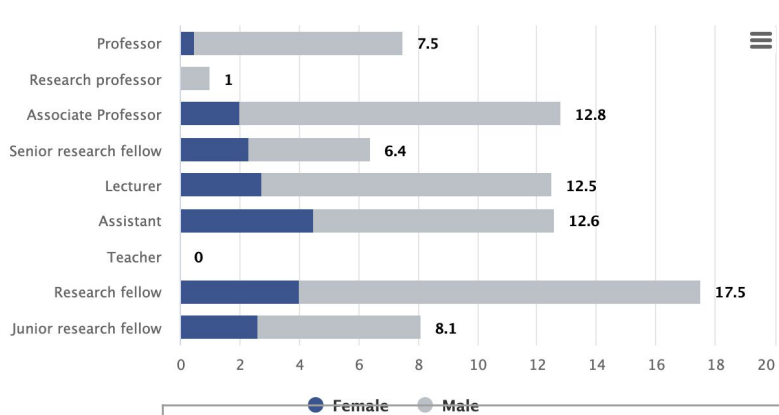
- Professor
- Associate Professor
- Research professor
- Associate Professor
- Senior research fellow
- Lecturer
- Junior Lecturer
- Assistant
- Teacher
- Research fellow
- Junior research fellow
- Professor emeritus



Gender breakdown of academic staff by position in 2022



Gender breakdown of academic staff by position in 2017

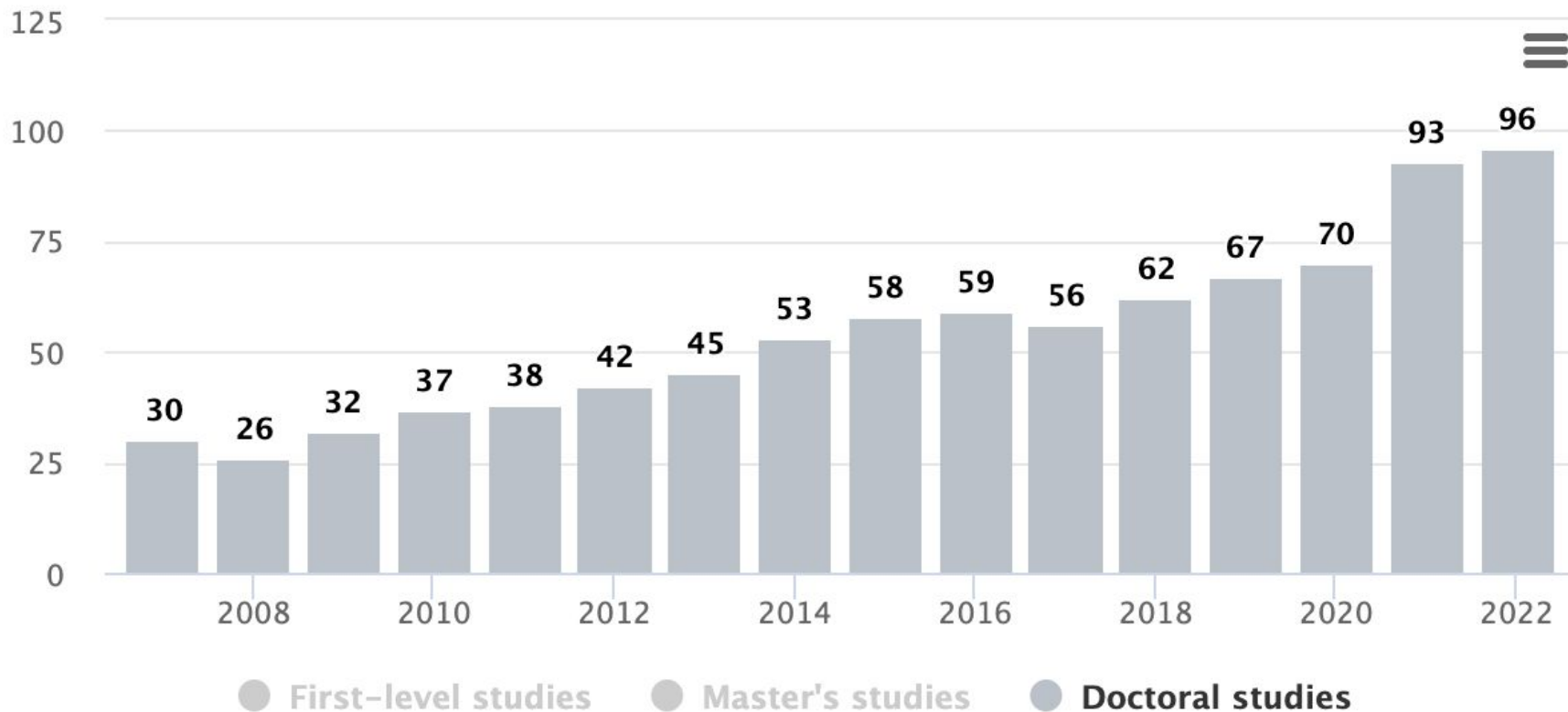


	2017 (end)	2022 (end)
<b>Professor</b>	8.5	12.3
<b>Assoc Prof, Sen. R. Fellow</b>	19.3	22.6
<b>Lecturer</b>	<b>12.5</b>	<b>32.15</b>
Research Fellow	17.5	21
Assistant, teacher, Jun. Lecturer	12.6	18.65
PhD students (enrolled)	<b>56</b>	<b>96</b>

# Number of students by level of study

Estonian students  International students

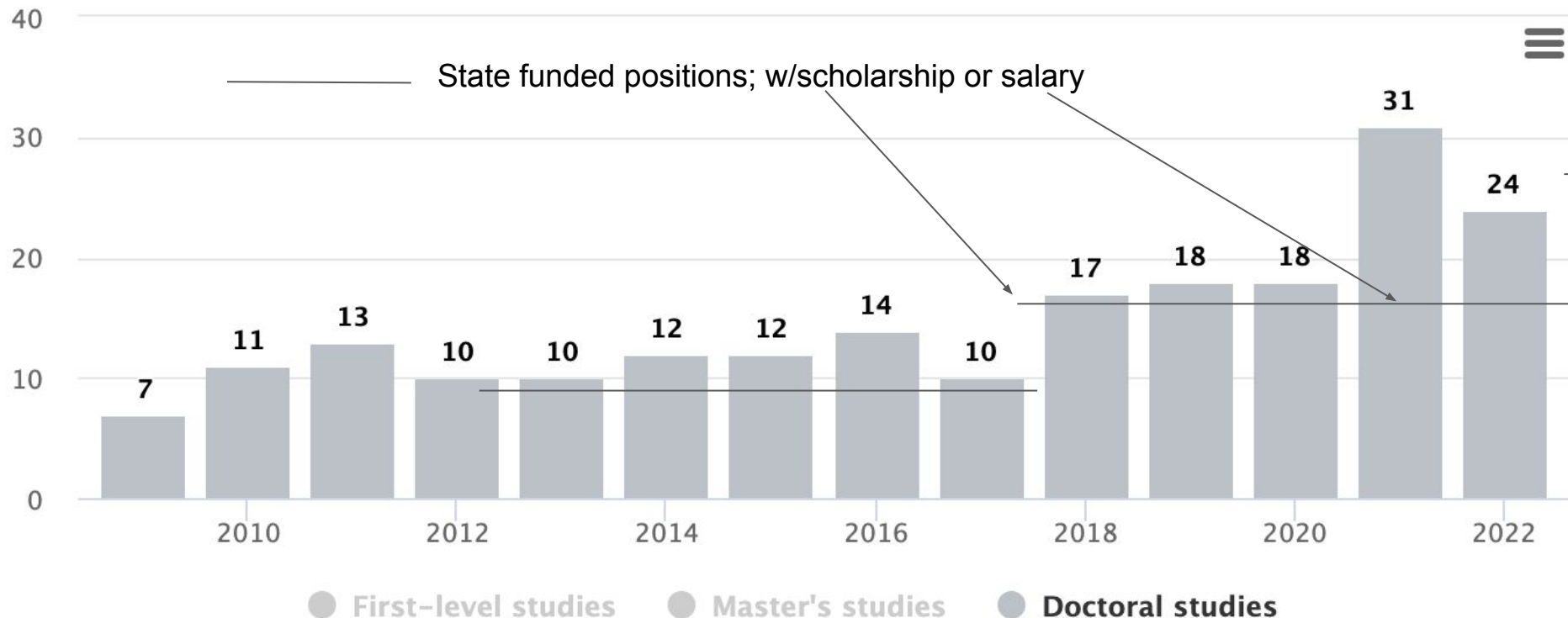
## PhD students



# Number of admitted students by level of study

# PhD admissions

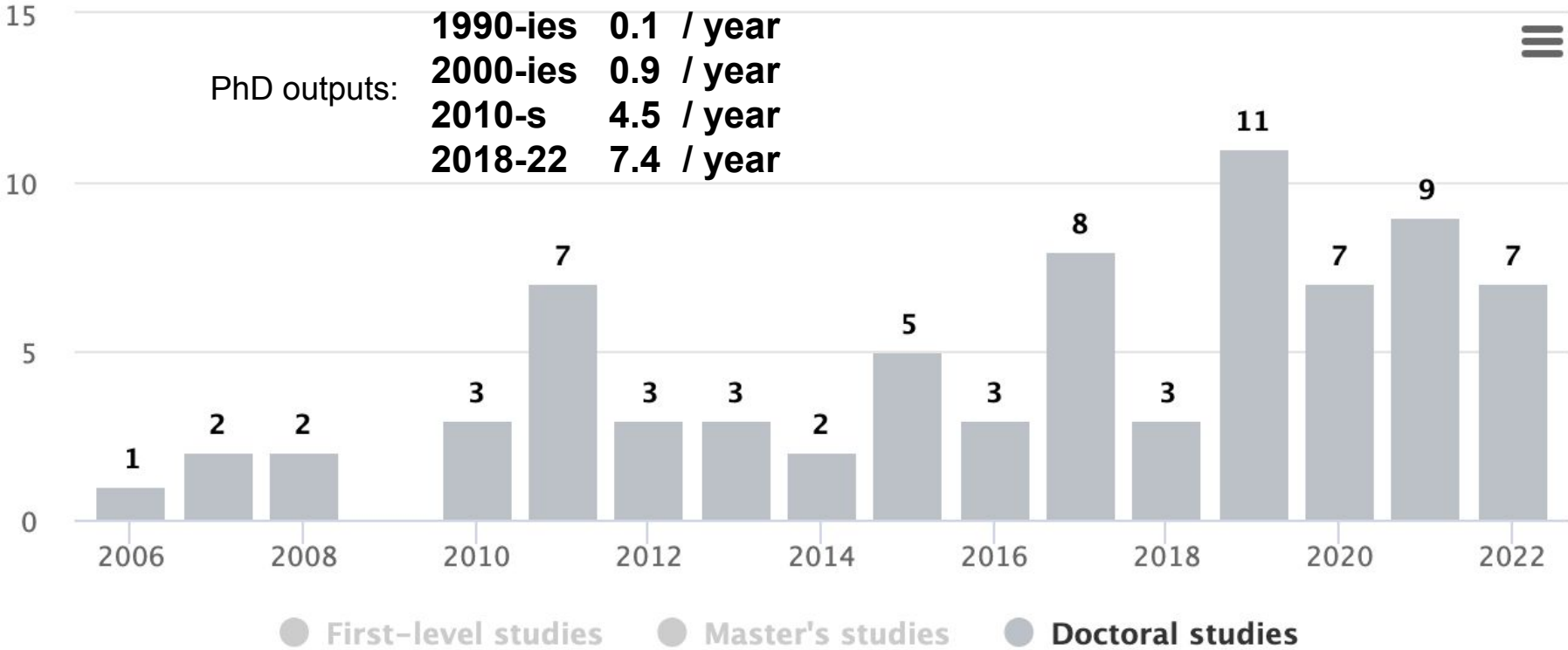
Estonian students  International students



# Number of graduates by level of study

Estonian students  International students

# PhD graduations



# Number of students by level of study

Estonian students  International students

## Estonian citizens (PhD students)

