

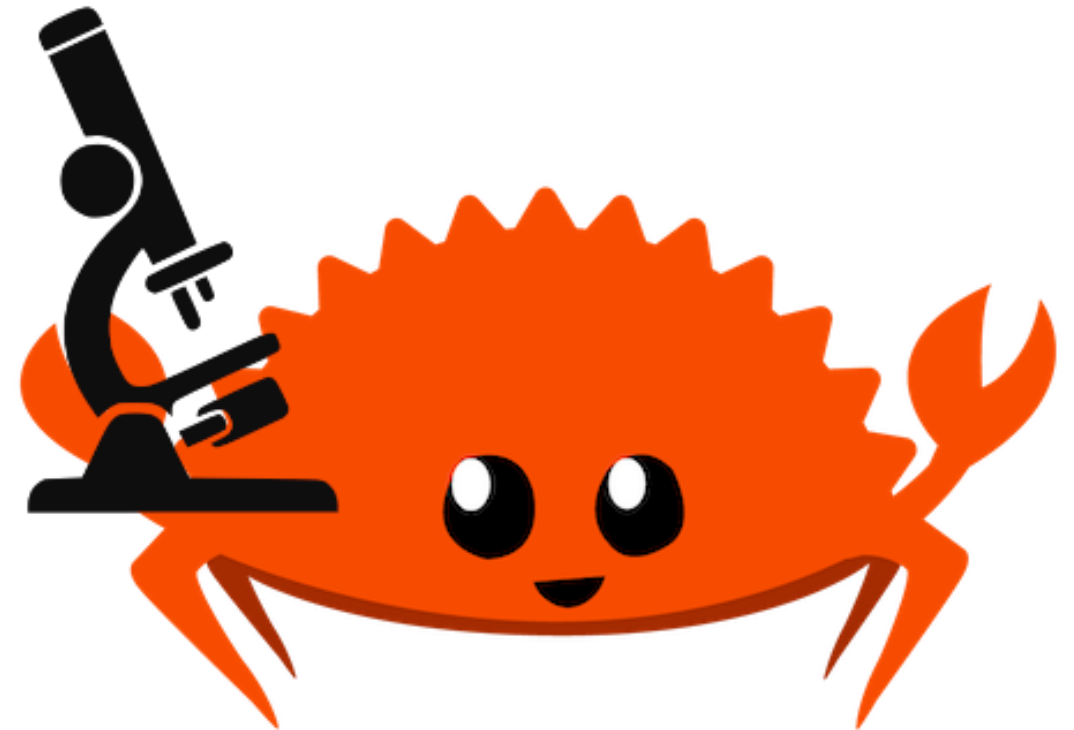


Fighting Cancer with Rust

Enola Knezevic

- 1 Federated Information Systems, German Cancer Research Center
- 2 Complex Data Processing in Medical Informatics, University Medical Center Mannheim

enola.knezevic@dkfz-heidelberg.de



Biobanks and data stores

- biospecimens, such as serum, plasma, tissue samples
- data about those samples and **pseudonymized** data about their patient donors
- researchers need to find samples and data by search criteria e.g.
 - sample type
 - storage temperature
 - molecular markers (nucleotide changes, amino-acid changes)
 - patient's diagnoses
 - therapies patients underwent

Full Parameter Search

Donor/Clinical Information

Gender

Diagnose ICD-10

equals: C61



Diagnosis age donor (years)

Date of diagnosis

Sample

Donor Age

Sample type

- Serum
- Tissue snap frozen
- Whole Blood
- Plasma
- Other derivative
- Other tissue storage
- Peripheral blood cells
- Urine
- RNA
- Other liquid biosample
- Buffy coat
- DNA
- Liquor/CSF
- Faeces
- Bone marrow
- Tissue (FFPE)
- Saliva
- Ascites
- Swab
- Dried whole blood

Sampling date

Storage temperature



Federated search results (BBMRI-ERIC)

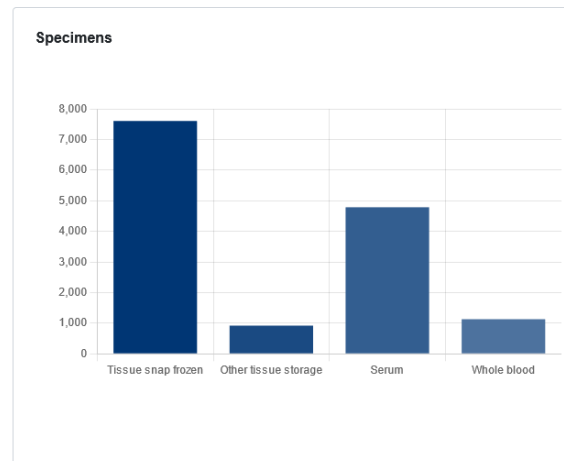
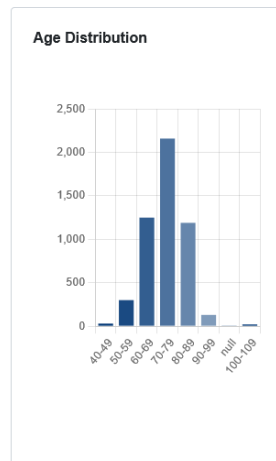
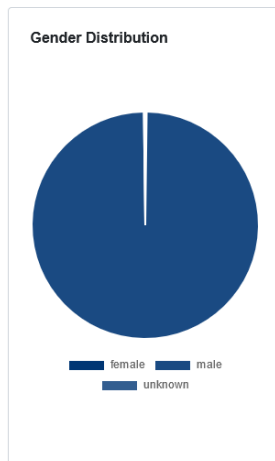
Results Sites: 15 Patients: 5050 Specimens: 14090

<input type="checkbox"/>	Sites	Patients	Specimen
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			
<input type="checkbox"/>			

<< < 1 2 > >>

Ask 0 sites to negotiate

- central search
- only receives **aggregated data** – individual data never leaves the site
- **obfuscated and rounded counts**



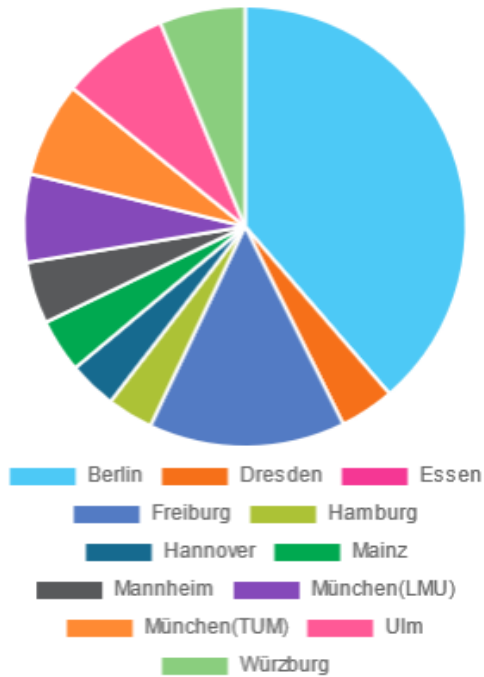
Federated search results (DKTK)

Ergebnisse

Standorte: 11 / 11

Patienten: 573335

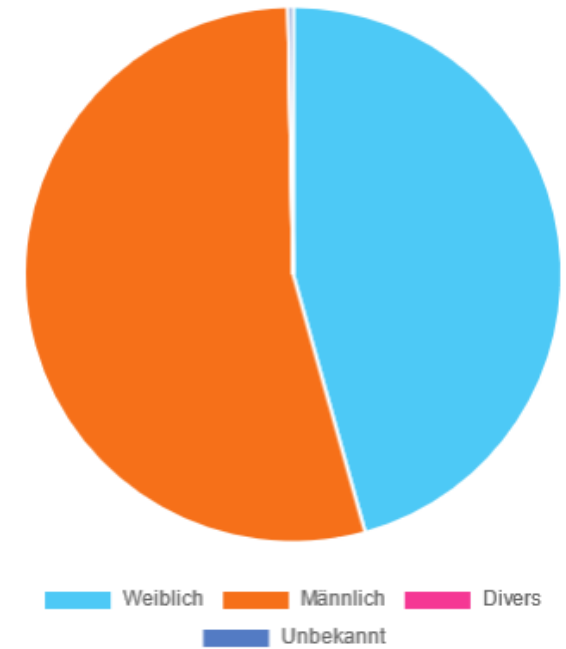
Patienten pro Standort



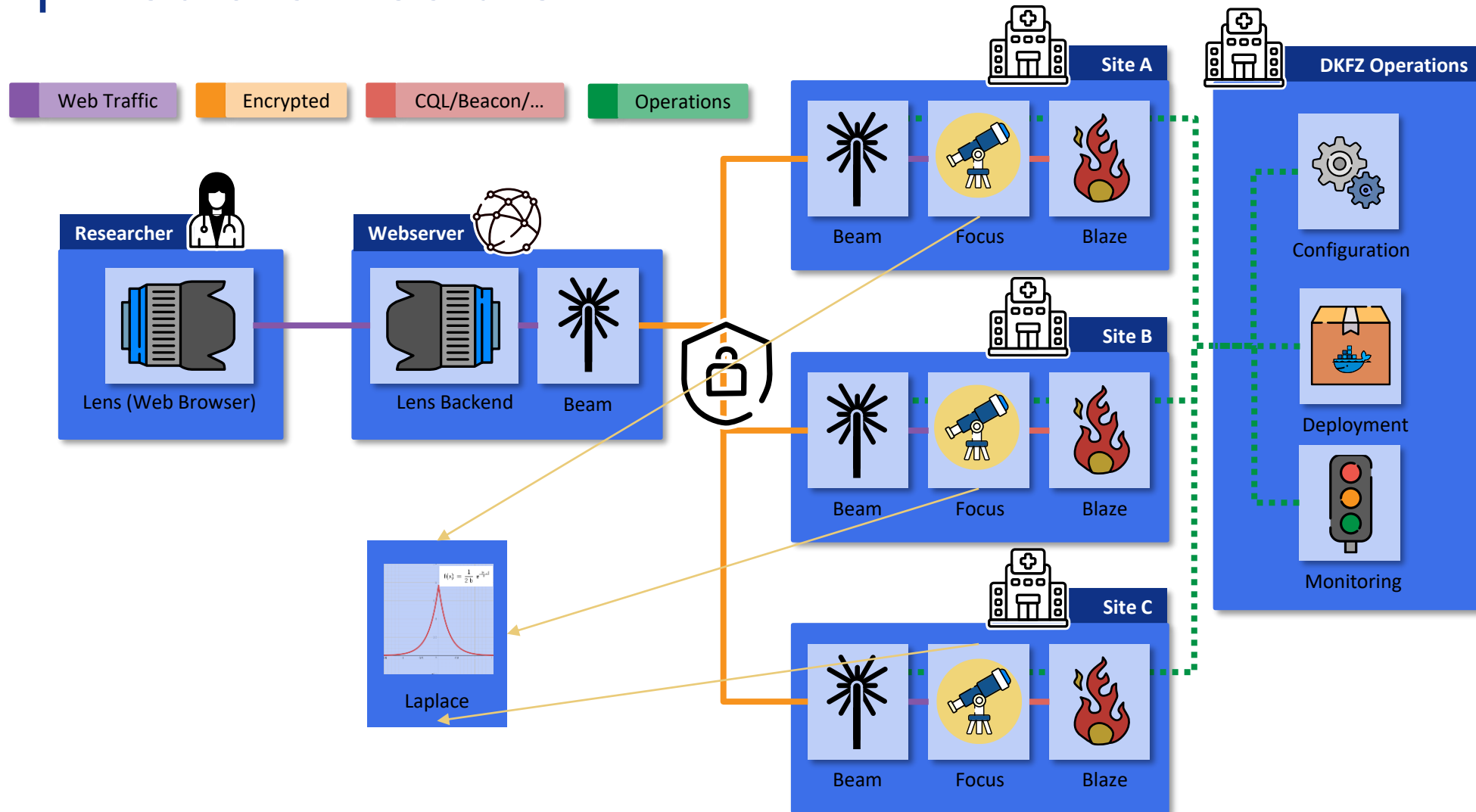
<input type="checkbox"/> Standorte	Patienten	Bioproben ⓘ
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		
<input type="checkbox"/>		

Navigation: ← 1 →

Geschlecht

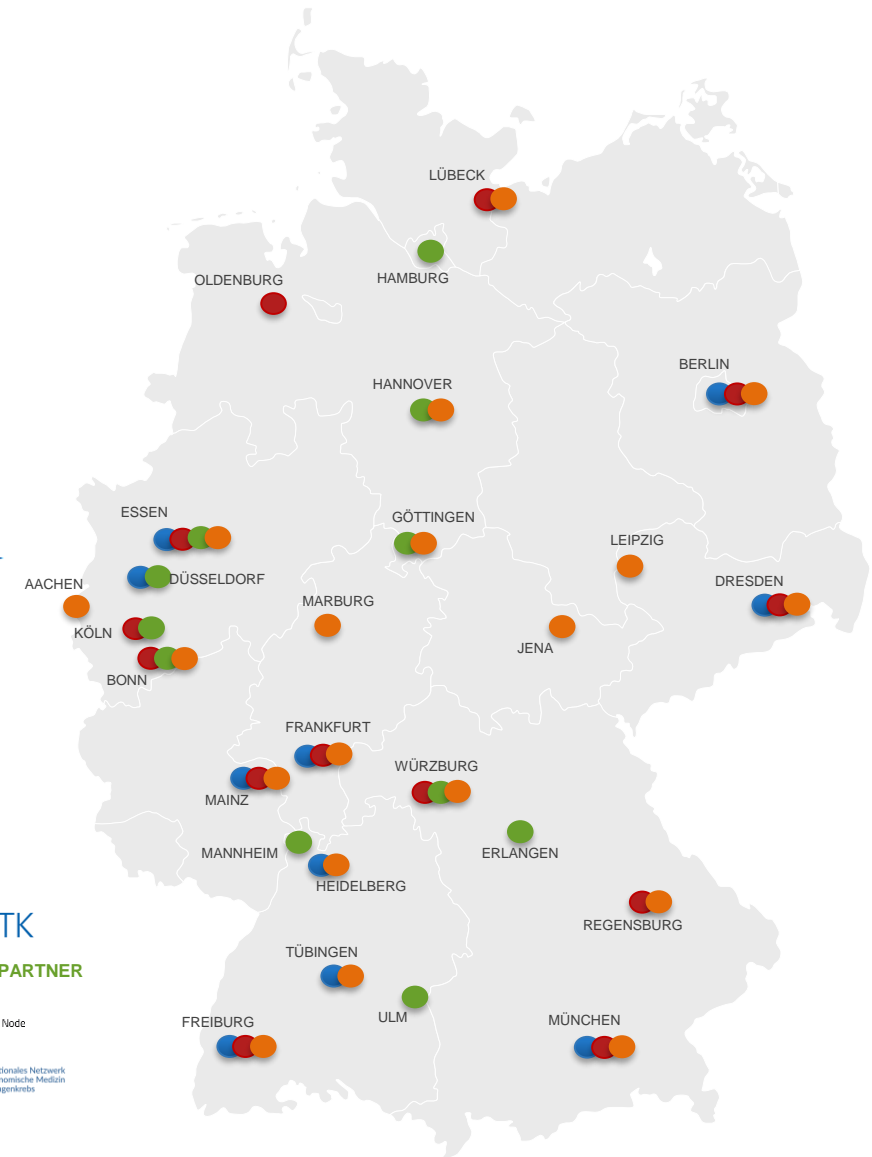


Simplified architecture



Projects

- GBN – German Biobank Node
- BBMRI-ERIC - Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium
- DKTK – German Cancer Consortium
- CCP – Clinical Communication Platform
- Cancer Core Europe
- ITCC P4 - Paediatric Preclinical Proof Of Concept Platform



EUCAIM
HOME PUBLIC CATALOGUE HELPDESK

▼ Patient

▼ Gender

male

female

other

unknown

> Age at Diagnosis

▼ Clinical Parameters

> Diagnosis

> Year of Diagnosis

▼ Image Parameters

▼ Modality

Magnetic Resonance Imaging

Positron Emission Tomography

Single Photon Emission Computed Tomography

Computed Tomography

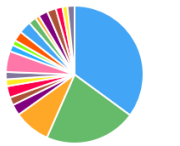
> Body Part

> Manufacturer

Results

Collections: 20 Patients: 21589


Studies per collection




■ UC1 ■ UC2
■ Lung cancer championship phase
■ Breast cancer MRI Only championship phase
■ Rectum cancer championship phase
■ Prostate cancer championship phase
■ Colon cancer championship phase
■ Prostate cancer classification phase

Collections	Provider	Studies	Subjects	
UC1	ProCancerI	8848	8826	▼
UC2	ProCancerI	5434	5432	▼
Lung cancer championship phase	CHAIMELEON	2149	816	▼
Breast cancer MRI Only championship phase	CHAIMELEON	650	332	▼
Rectum cancer championship phase	CHAIMELEON	475	313	▼
Prostate cancer championship phase	CHAIMELEON	681	677	▼
Colon cancer championship phase	CHAIMELEON	408	396	▼
Prostate cancer classification phase	CHAIMELEON	433	431	▼
Lung cancer classification phase	CHAIMELEON	1239	482	▼
Lung Cancer Only Images (July 23)	CHAIMELEON	401	271	▼
Prostate Cases MRI Only (July 23)	CHAIMELEON	306	298	▼
Rectum Cancer MRI Only (July 23)	CHAIMELEON	583	429	▼
Colon Cancer CT Only (July 23)	CHAIMELEON	734	668	▼
Breast Cancer MRI_only (July 23)	CHAIMELEON	456	256	▼
Breast Cancer Only Images v2	CHAIMELEON	239	110	▼
Rectum Cancer Only Images v2	CHAIMELEON	551	403	▼
Colon Cancer Only Images v2	CHAIMELEON	528	468	▼
Lung Cancer Only Images v2	CHAIMELEON	397	269	▼


This federated search was made with the open source [Samplify tools \(Lens, Beam, Focus, Bridgehead\)](#), created by the [German Cancer Research Center \(DKFZ\)](#).



[PRIVACY POLICY](#) [COOKIES POLICY](#)




UNIVERSITÄTSMEDIZIN
MANNHEIM



GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

Medizinische Fakultät Mannheim
der Universität Heidelberg
Universitätsklinikum Mannheim



Samply.Beam

- Distributed task broker designed for efficient communication across strict network environments present in medical informatics:
 - End-to-end encryption
 - Certificate management and validation
 - Only outbound connections



Samplify.Focus



- Federated query dispatcher working with Samplify.Beam
- CQL query generation to prevent CQL injections
- AST translation: EUCAIM: Chameleon, ProCAncer-I
- Running the query against the data stores, other applications
- Query result obfuscation using Samplify.Laplace library



Differential privacy algorithms

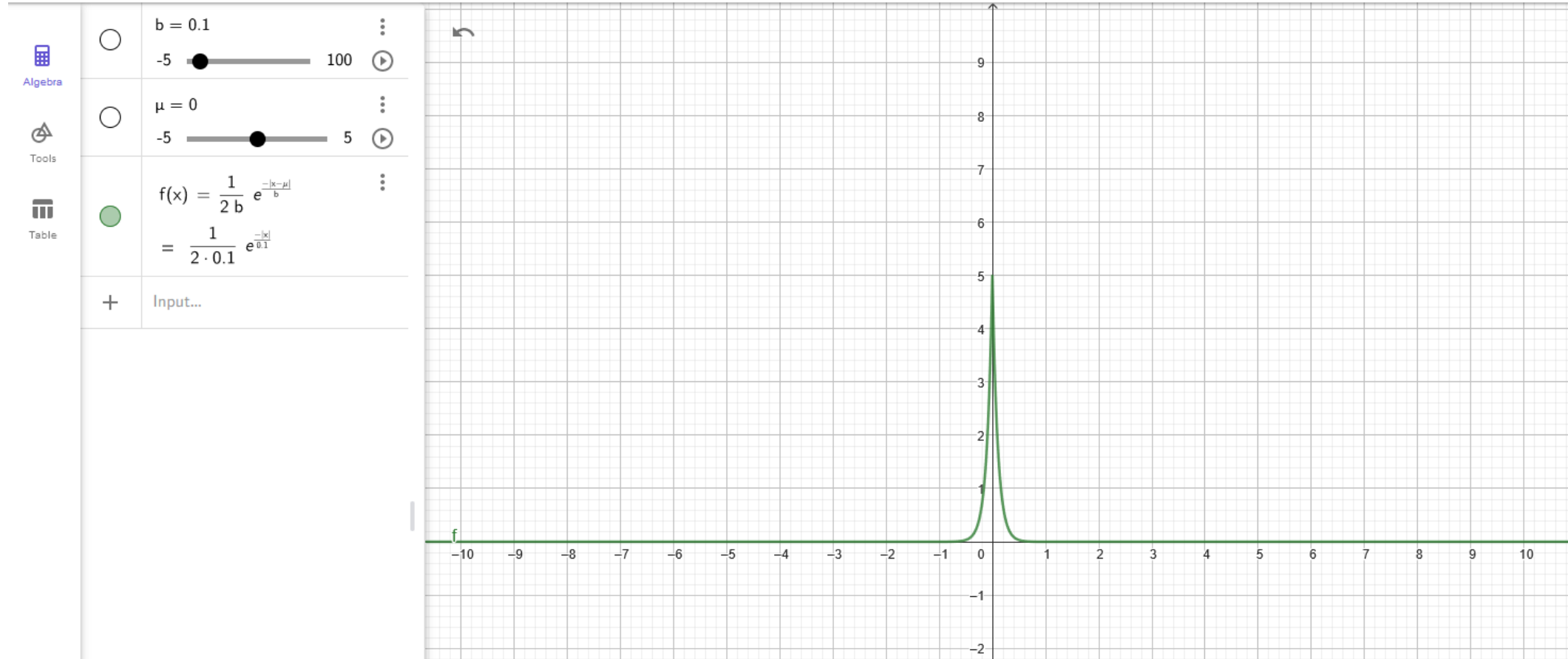
- Promises:
 - Preserve privacy while keeping the data useful for research: publish aggregate data, withhold individual data
 - Resist differencing attacks trying to identify whether an individual's data is in a certain database or not (e.g. by selecting a certain diagnosis, date of diagnosis, age, and gender of the patient)
 - **Offer similar level of privacy as having individual's data removed from the database**

Why k-anonymity is not enough

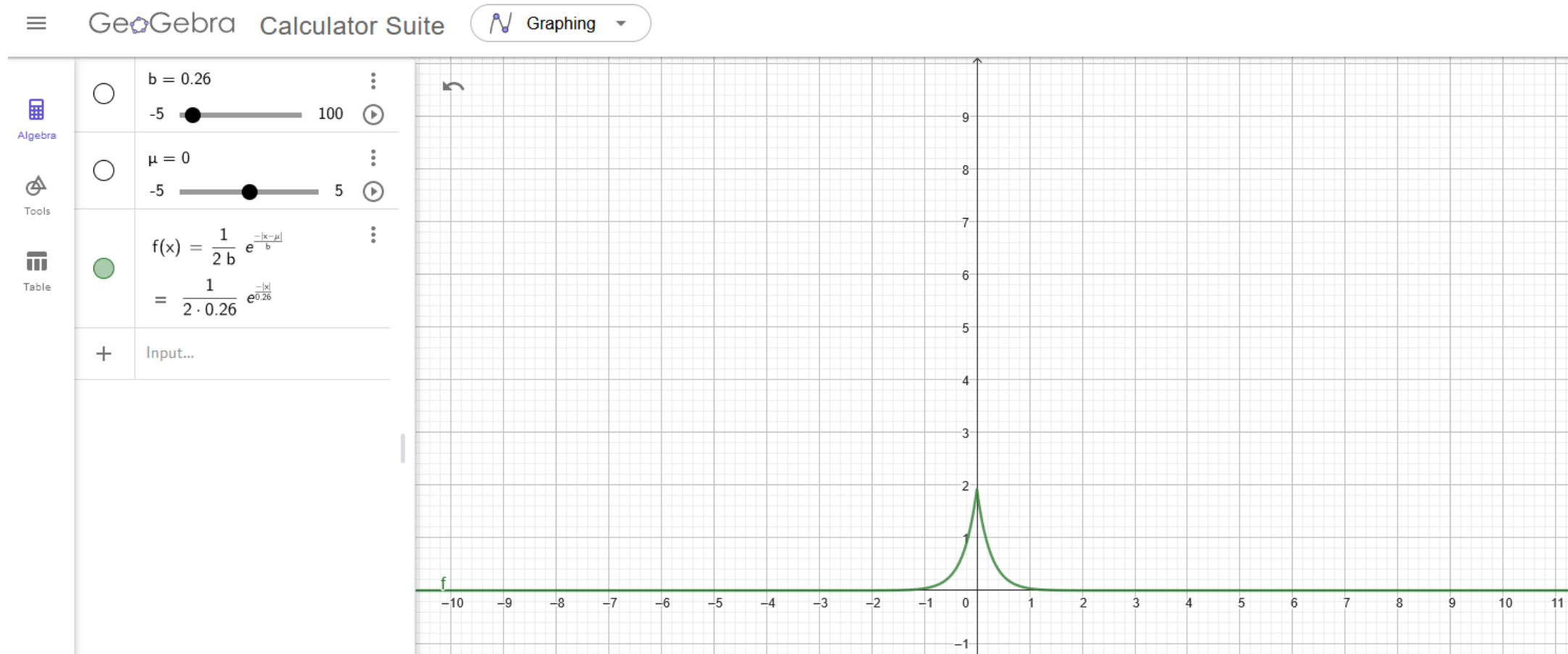
- Medical data often anonymised and pseudonymised
- k-anonymity - all combinations of attributes are satisfied by at least k entries in the dataset
- “Cut-off” value of k not sufficient to ensure k-anonymity, especially if multiple sensitive criteria are involved
- Sensitive characteristic evenly distributed within a class - can be inferred
- Associating anonymised information with additional external information

Laplace distribution

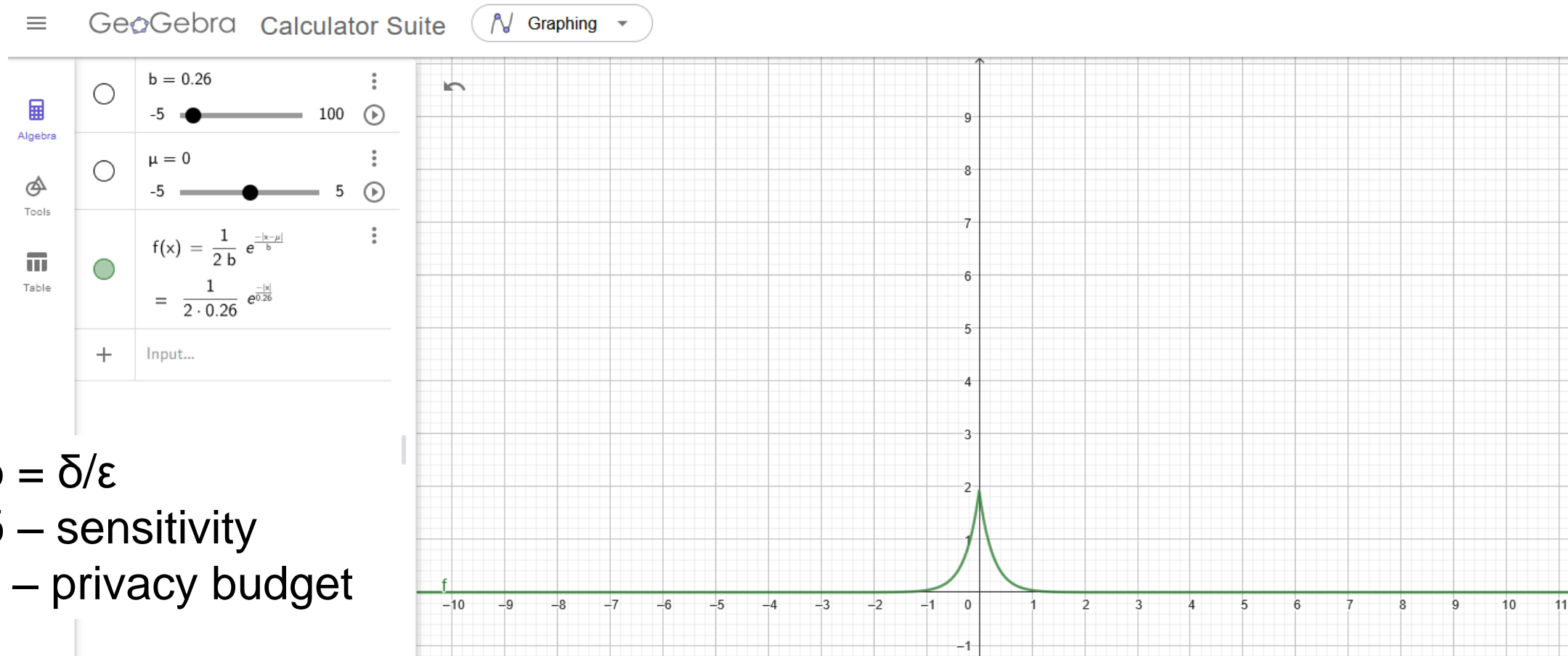
GeoGebra Calculator Suite Graphing



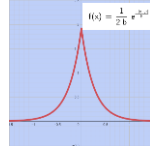
Privacy – usability trade-off



Privacy – usability trade-off



Sampl.Laplace



- differential privacy-inspired query result obfuscation
- Rust crate (& a Java library in case anyone prefers that)
- highly configurable:
 - sensitivity, privacy budget
 - values under 10: change all to 10, change all to 0, or obfuscate in the usual way
 - turn off obfuscation of zeroes
 - rounding step
 - obfuscation value cache – consistent results



Open-source

- Apache-2.0 license
- Feel free to use and contribute to our software



Join us

