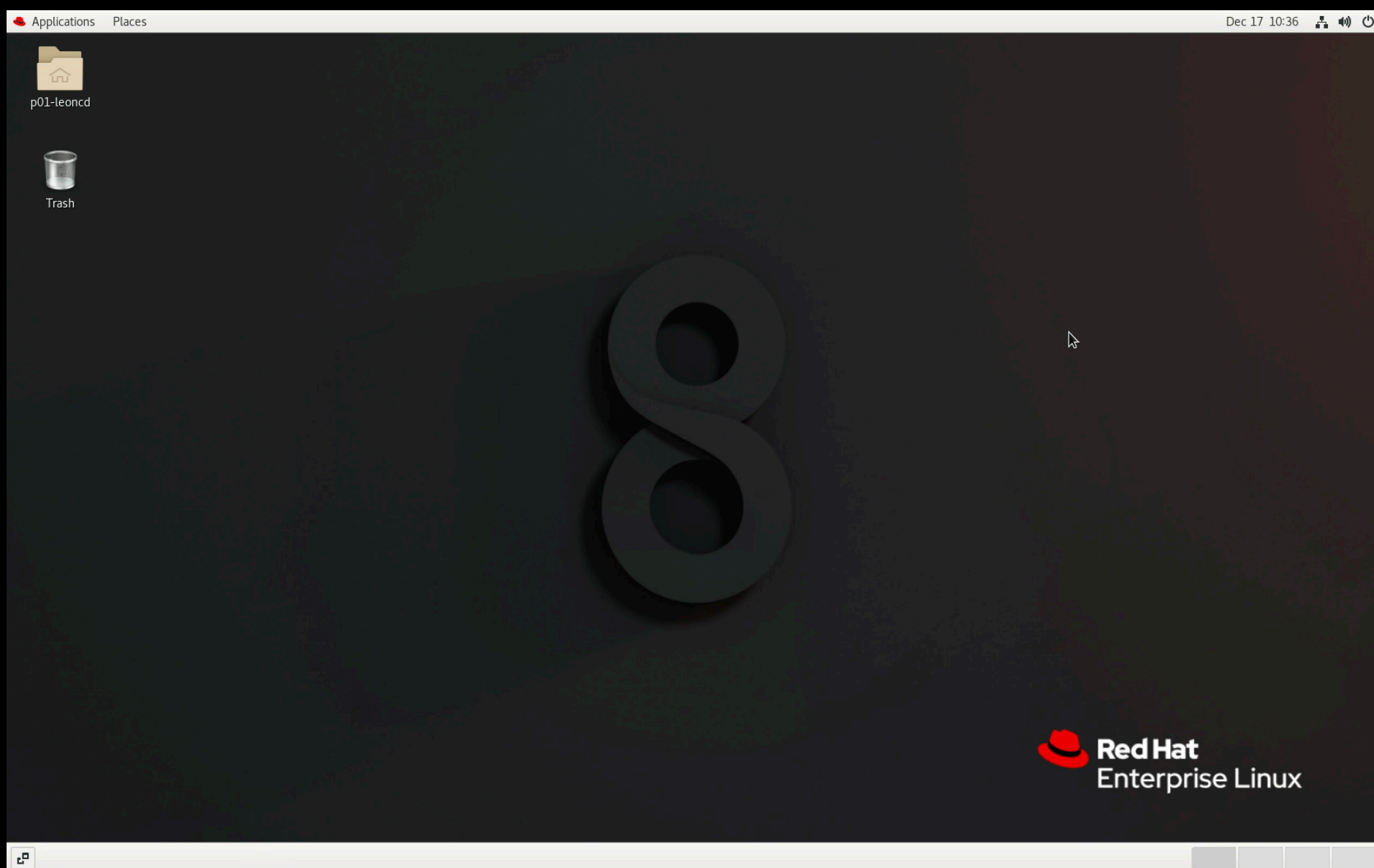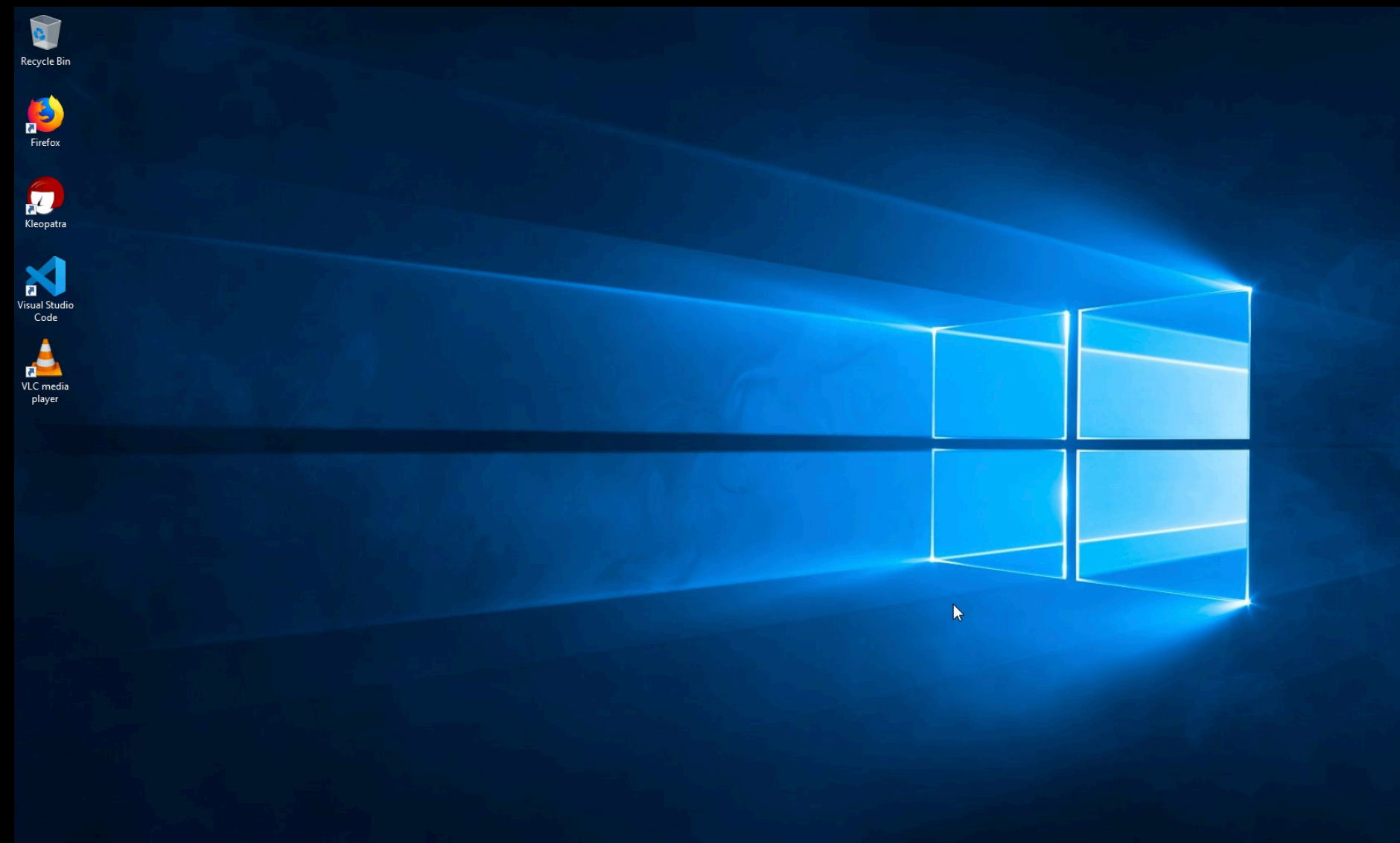# TSD: 2020 - 2022

What we've been doing to make research with sensitive data easier

Leon du Toit - Section Leader, Research Services, USIT

What is TSD?

- Secure cloud, bring-your-own:

  - Ethical approval

  - Data

  - Software

  - License

- Remote login, from anywhere in the world

- Platform for app development

```
ldt:~ leondutoit$ tacl --help
Usage: tacl [OPTIONS] [PNUM]

  tacl - TSD API client.

Options:
  --guide TEXT               Print a guide
  --env TEXT                 API environment  [default: prod]
  --group TEXT               Choose which file group should own the data
                             import
  --basic                    To use basic auth
  --upload TEXT              Import a file or a directory located at given
                             path
  --upload-id TEXT           Identifies a specific resumable upload
  --resume-list              List all resumable uploads
  --resume-delete TEXT       Delete a specific resumable upload
  --resume-delete-all        Delete all resumable uploads
  --download TEXT            Download a file
  --download-list            List files available for download
  --download-id TEXT         Identifies a download which can be resumed
  --version                  Show tacl version info
  --verbose                  Run tacl in verbose mode
  --config-show              Show tacl config
  --config-delete            Delete tacl config
  --session-delete           Delete current tacl login session
  --register                 Register tacl for a specific TSD project and
                             API environment
  --ignore-prefixes TEXT     Comma separated list of sub folders to ignore
                             (based on prefix match)
  --ignore-suffixes TEXT     Comma separated list of files (based on suffix
                             match)
  --upload-cache-show        View the request cache
  --upload-cache-delete TEXT Delete a request cache for a given key
  --upload-cache-delete-all  Delete the entire request cache
  --cache-disable            Disable caching for the operation
  --download-cache-show      View the request cache
  --download-cache-delete TEXT Delete a request cache for a given key
  --download-cache-delete-all Delete the entire request cache
  --upload-sync TEXT         Sync a local directory, incrementally
  --download-sync TEXT       Sync a remote directory, incrementally
  --cache-sync               Enable caching for sync
  --keep-missing             Do not delete missing files in the target
                             directory while syncing
  --keep-updated             Do not over-write updated files in the target
                             directory while syncing
  --download-delete TEXT     Delete a file/folder which is available for
                             download
  --api-key TEXT             Pass an explicit API key - over-rides tacl
                             config
  --help                     Show this message and exit.
ldt:~ leondutoit$
```

Nettskjema Bilde Cancel

Posisjon

Lengdegrader:	10.478162
Breddegrader:	59.880078
Høyde:	50 m.o.h.
Hastighet:	0.6 m/s
Nøyaktighet:	14.2 meter

Tidsstempel:
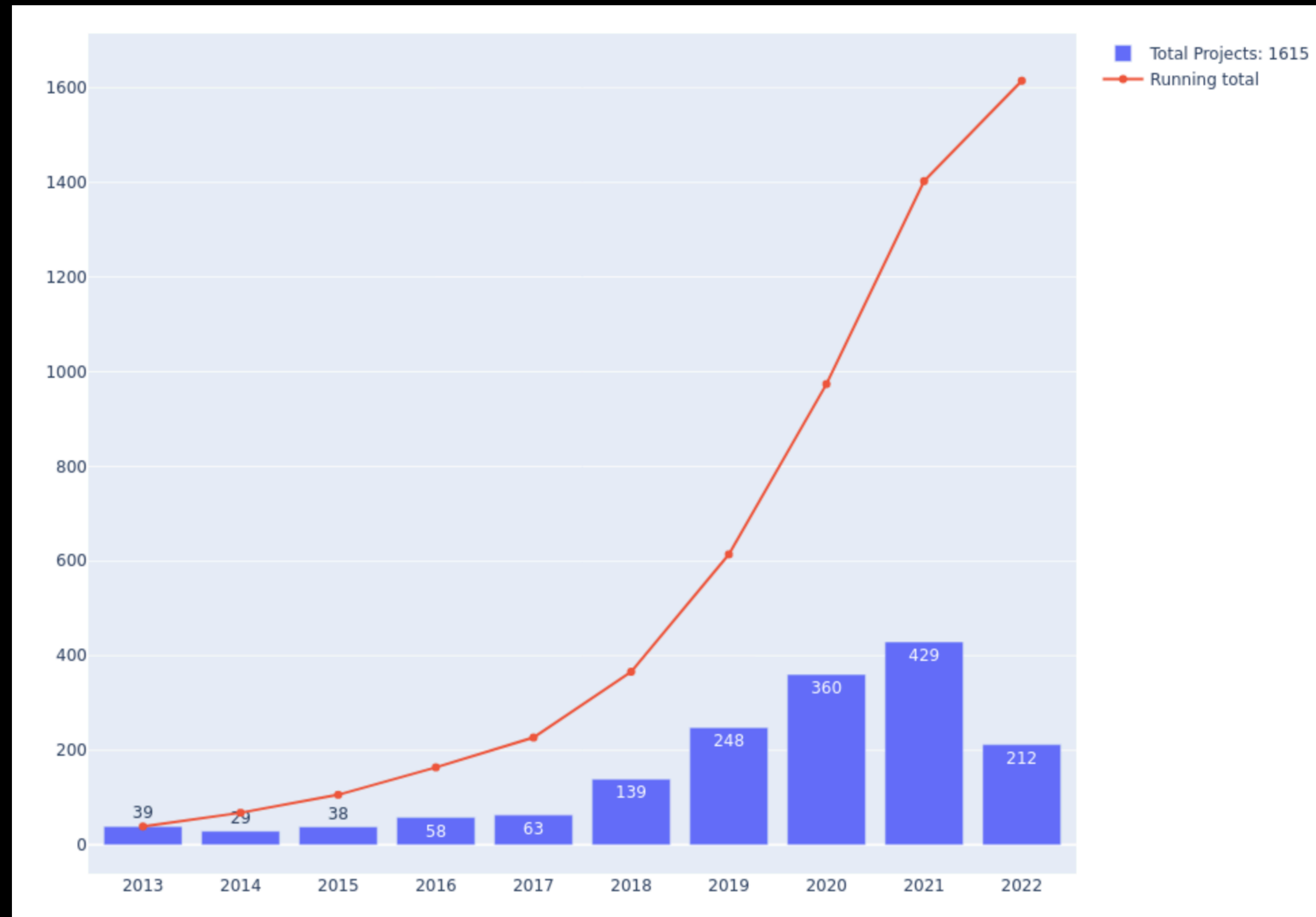
12:40 26 August 2020

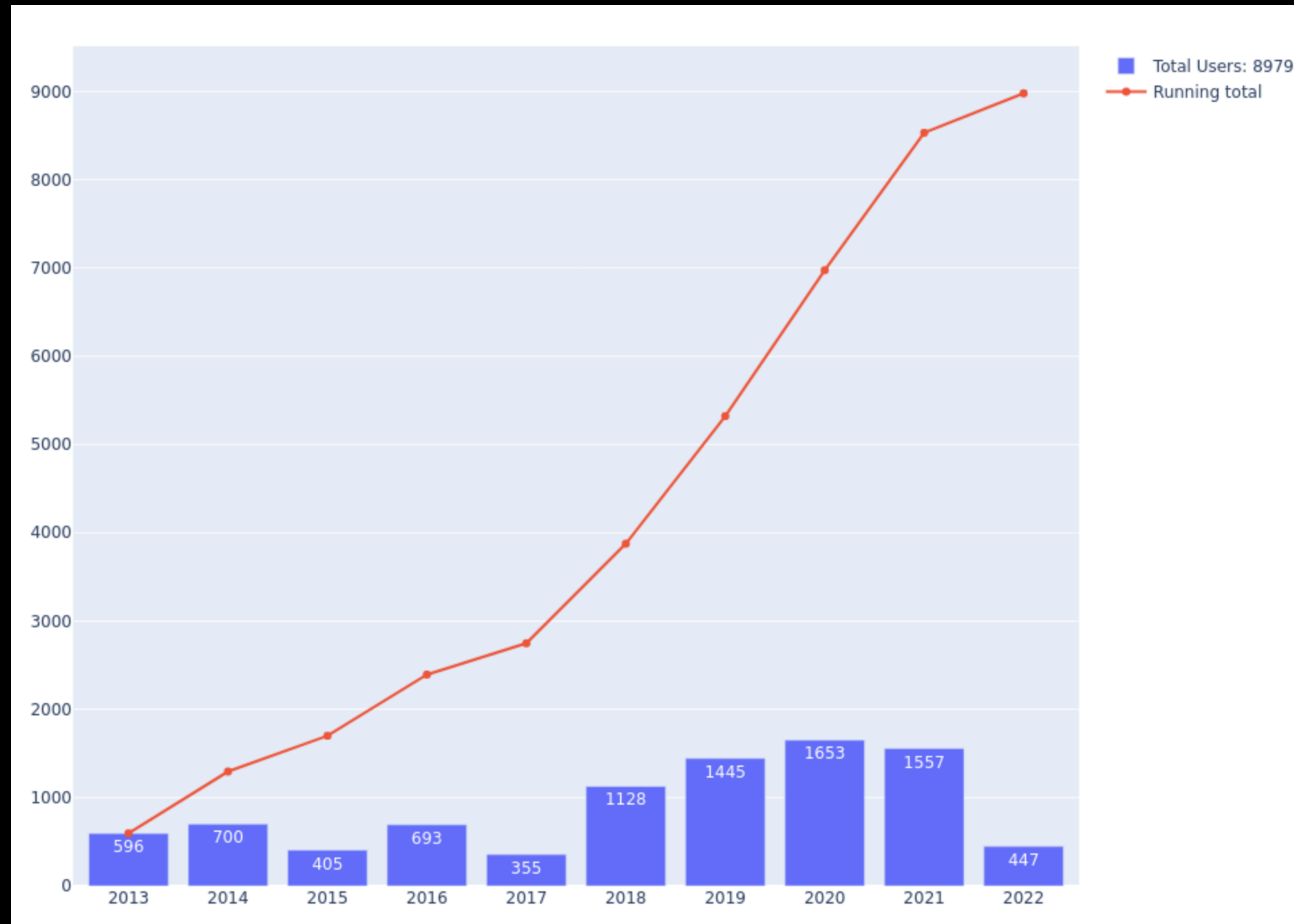**Nøyaktighet: 14 meter**

Lagre posisjon

Context - growth and security

Growth
Projects - more than one a day

Growth
Users

Growth
Data volumes and traffic

- Storage (2020 - current)

  - 7PB - 10.4PB

- Web traffic

  - Up to 1 million requests per day

Growth
Institutional agreements

- 36 institutional commercial and data processing agreements in total

- 19 new since 2020

- OsloMet, NUBU, AHUS, FHI, BI, SINTEF, NORCE, St. Olavs Hospital, NKVTS

Security
Threats

- War in Ukraine

- Ransomware-as-a-service

- Bugs in IT systems

Security
Trends

- Small attacks often the start of bigger attacks

- Advanced persistent attacks now a real possibility

- Suspect that a ransomware group has tried to attack UiO

Security
Responses

- More early warning systems

- Two-factor login on all services

- TSD builds on and relies on University sector and UiO's security work, e.g.:

  - network scanning

  - early warning systems

A hierarchy of needs

| Share data | Build an app | Export publication material |

Share data	Build an app	Export publication material

Process data	Apply access control

Share data	Build an app	Export publication material

Process data	Apply access control

Install software	Store data	Compute

Share data	Build an app	Export publication material
Process data		Apply access control
Install software	Store data	Compute
Import/collect data		Login

| Share data | Build an app | Export publication material |

| Process data | Apply access control |

| Install software | Store data | Compute |

| Import/collect data | Login |

| User account | TSD project |

Share data	Build an app	Export publication material

Process data	Apply access control

Install software	Store data	Compute

Import/collect data	Login

User account	TSD project

Agreements and approvals

Share data	Build an app	Export publication material

Process data	Apply access control

Install software	Store data	Compute

Import/collect data	Login

User account	TSD project

Agreements and approvals

Stable
Correct
Confidentiality
Support

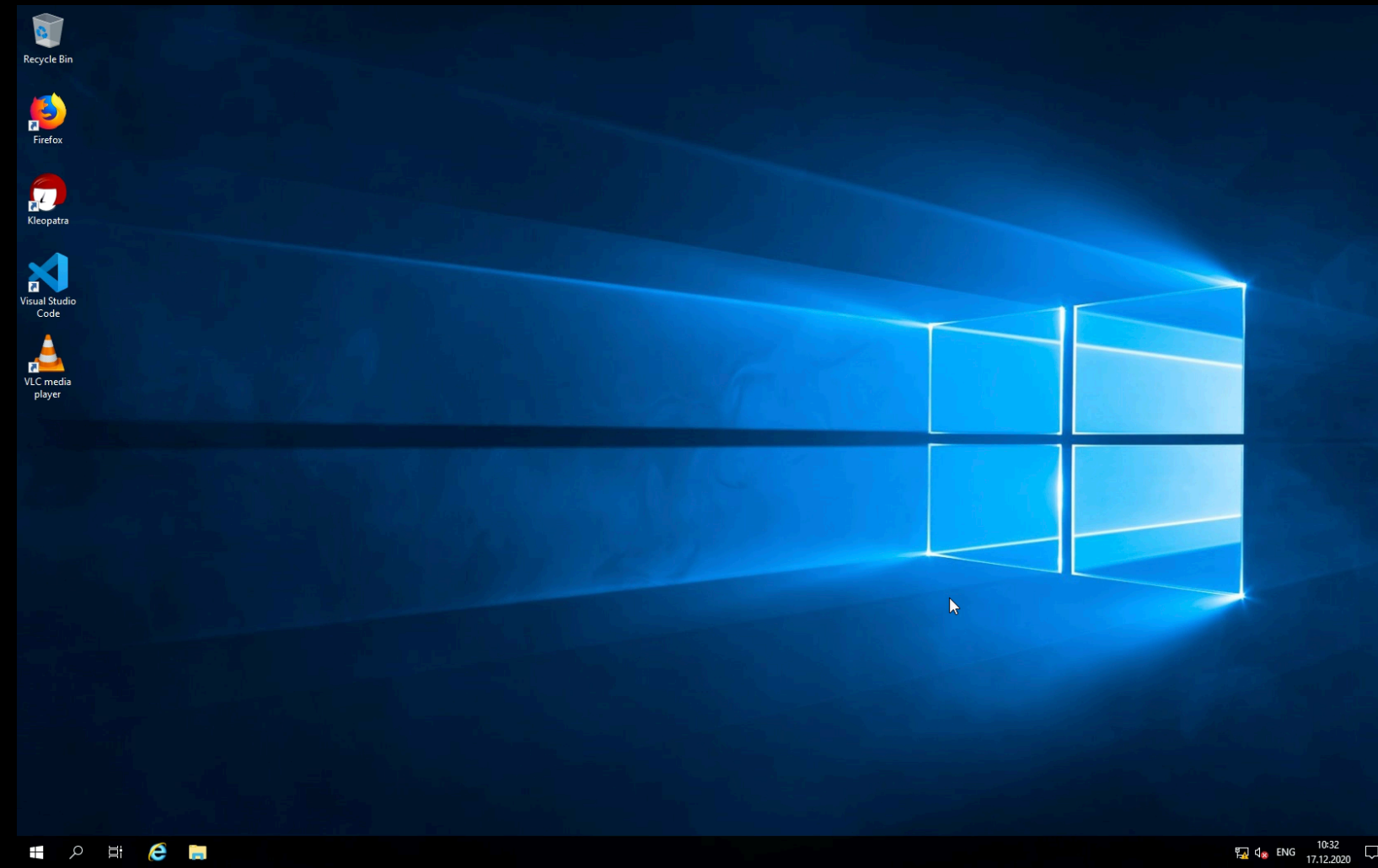# Stability
Login

```
#define PF_ANCHOR_STACKSIZE        32
struct pf_kanchor_stackframe {
        struct pf_kruleset      *rs;
        struct pf_krule         *r;      /* XXX: + match bit */
        struct pf_kanchor       *child;
};
```

Firewall

- Pre-2020: recurrent issues - operating system deadlock and crash

- Very problematic: nothing works

- Long road to a solution

  - New machines, new network cards (advice from Netflix engineers)

  - Hot failover - if one crashes, another takes over

  - Engaged operating system developers - found source of the bug

  - Re-write firewall rules to avoid problematic code branch

  - Establish test environment, performance tuning, stabilise operations
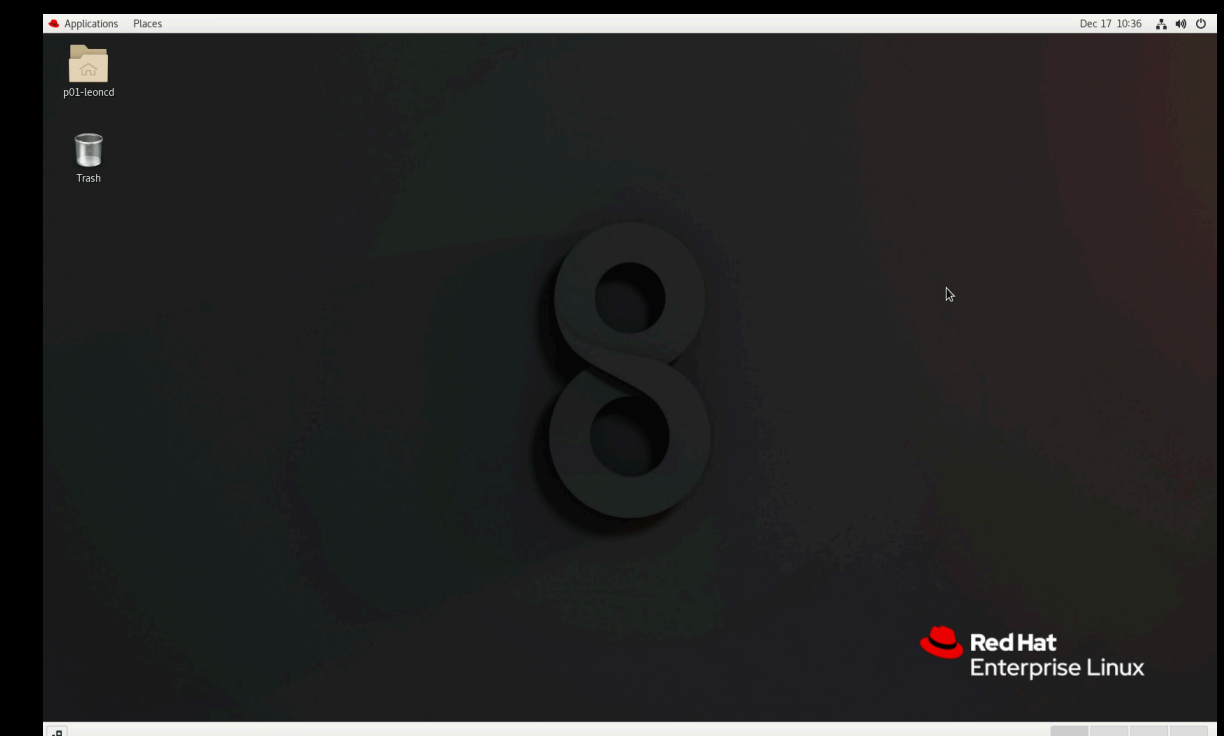
Virtual machines
hardware, configuration

- Invest heavily into efficient operations

  - Standardised installation of components across the organisation

  - Team-based knowledge bases

  - Enforce hardware standards: Intel CPUs, Nvidia GPUs

Modernising linux
Redhat 8

- Single login method - VMware Horizon

- Builds upon aforementioned operational stability

- More modern operating system enables easier software management

- Migrated hundreds of projects with custom software

- More than a year of work

Correctness
Data transport

Therapy training
Recoring sessions with web-cams

- Web-based, stream video and audio to TSD, analyse video

- 2022:

    - 150 students per semester

    - 1500 hours of video recorded

    - 1.5TB of data

- Cannot lose data while sending it to TSD

- Using internet: we do not have full control over the network

Confidentiality
Advanced access control

Forskningsprosjekter ved Avdeling for Ernæringsvitenskap

DIGIKOST: Digital registration of dietary intake and other lifestyle factors

FoodCapture

Løsning for sykdomsrelatert underernæring, til sykehus, sykehjem og hjemmetjeneste

Challenge

- Share data with a specific person

- Without…

  - Making them a member of your project

  - Login to a virtual machine

TSD Project roles

- Principal Investigator

- Administrator

- Member

- 2020: All of which required a user which can login
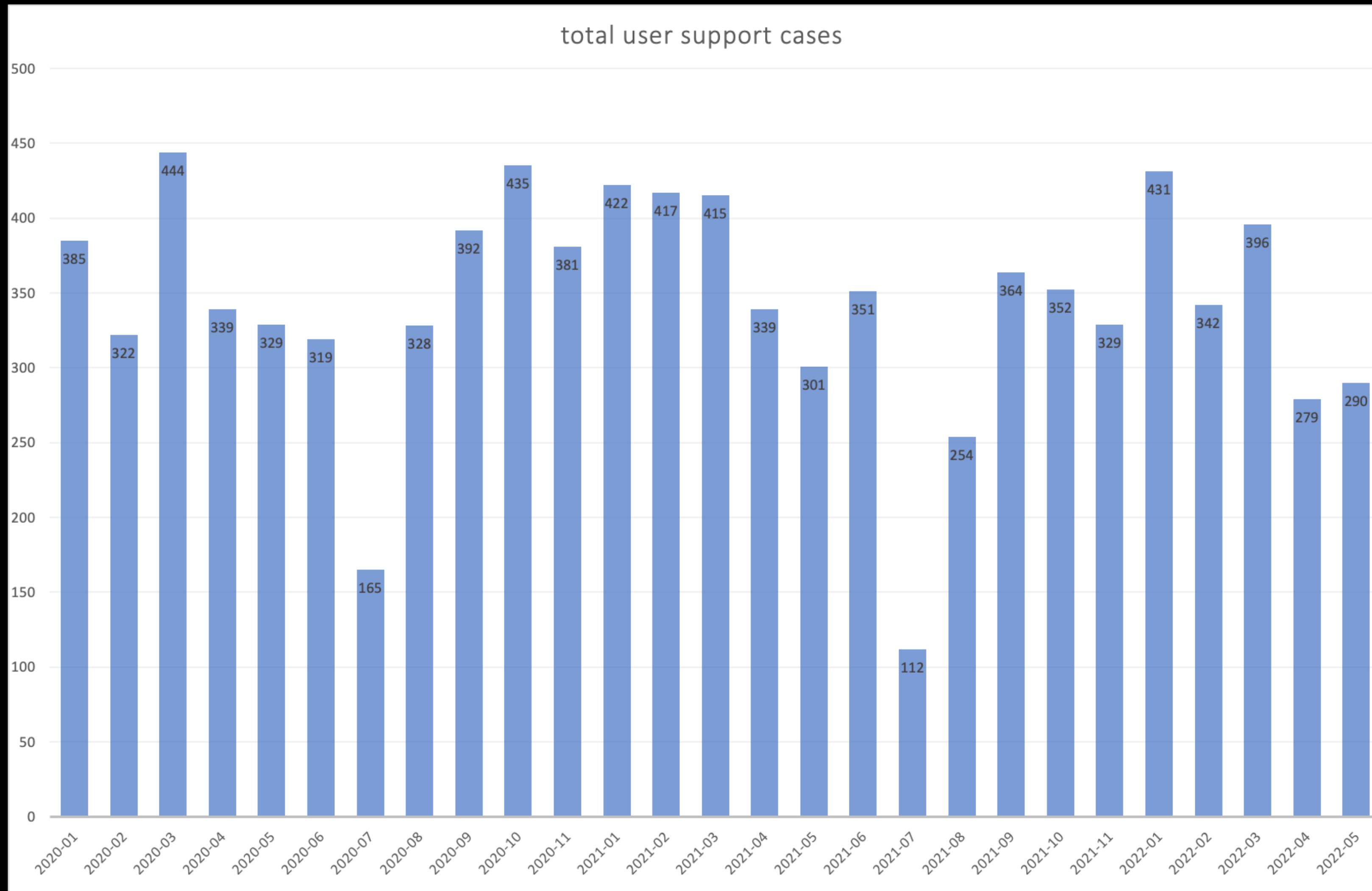
Solution?
Build a new identity management (IAM) system

- New role:

  - Project associate

  - Access to specific data only

- 2 years of work: seamless adoption

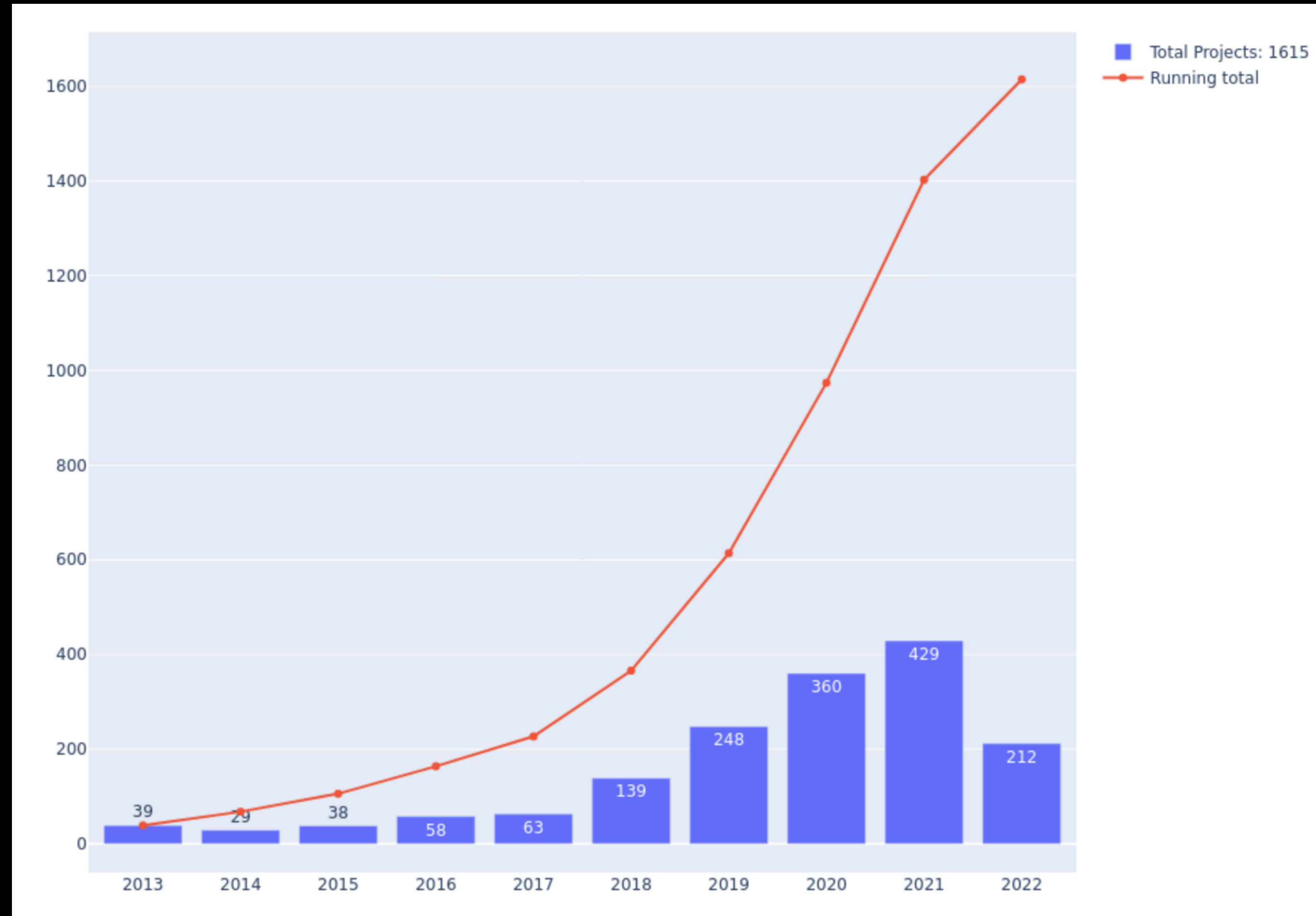- Foundation for self-service

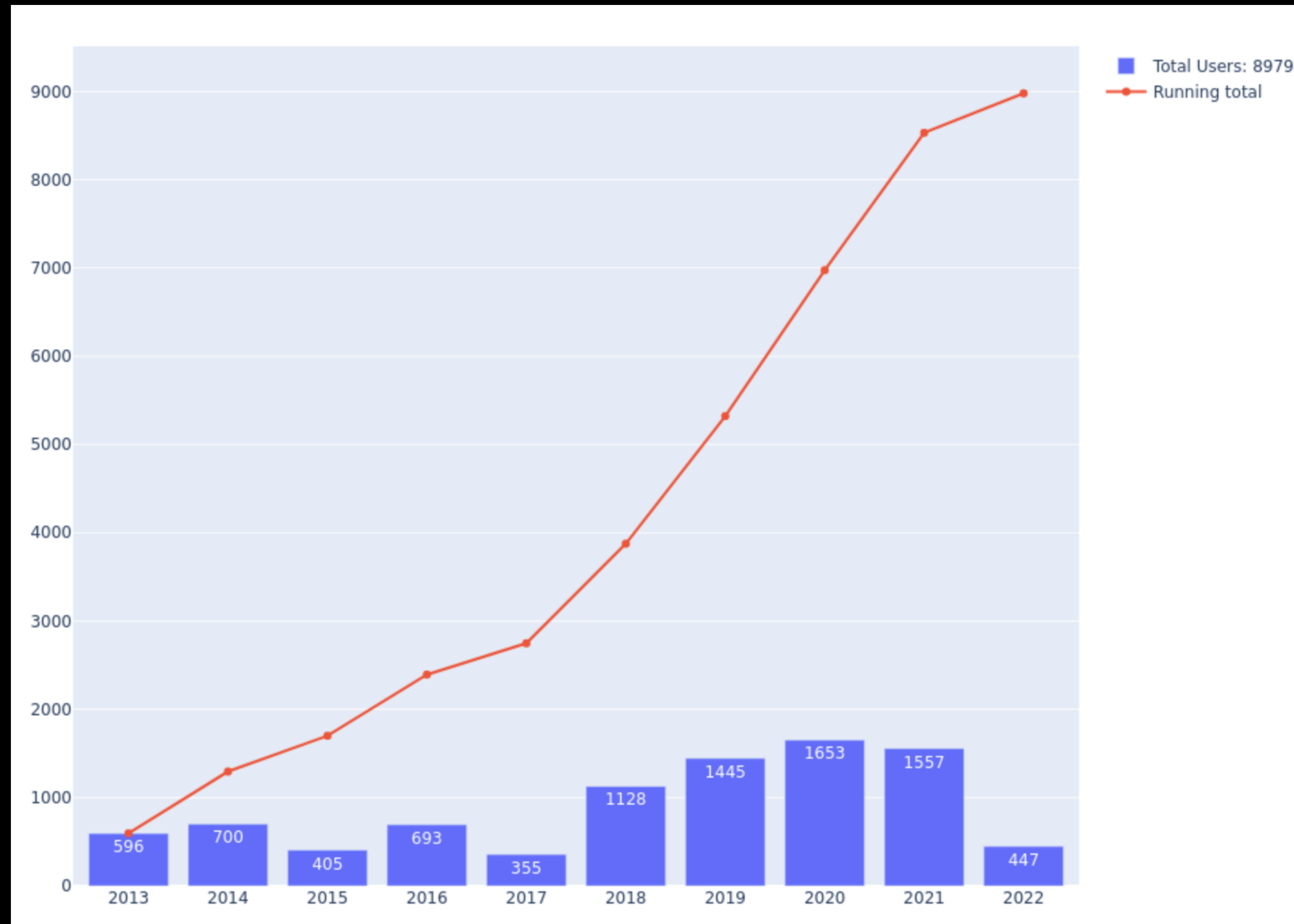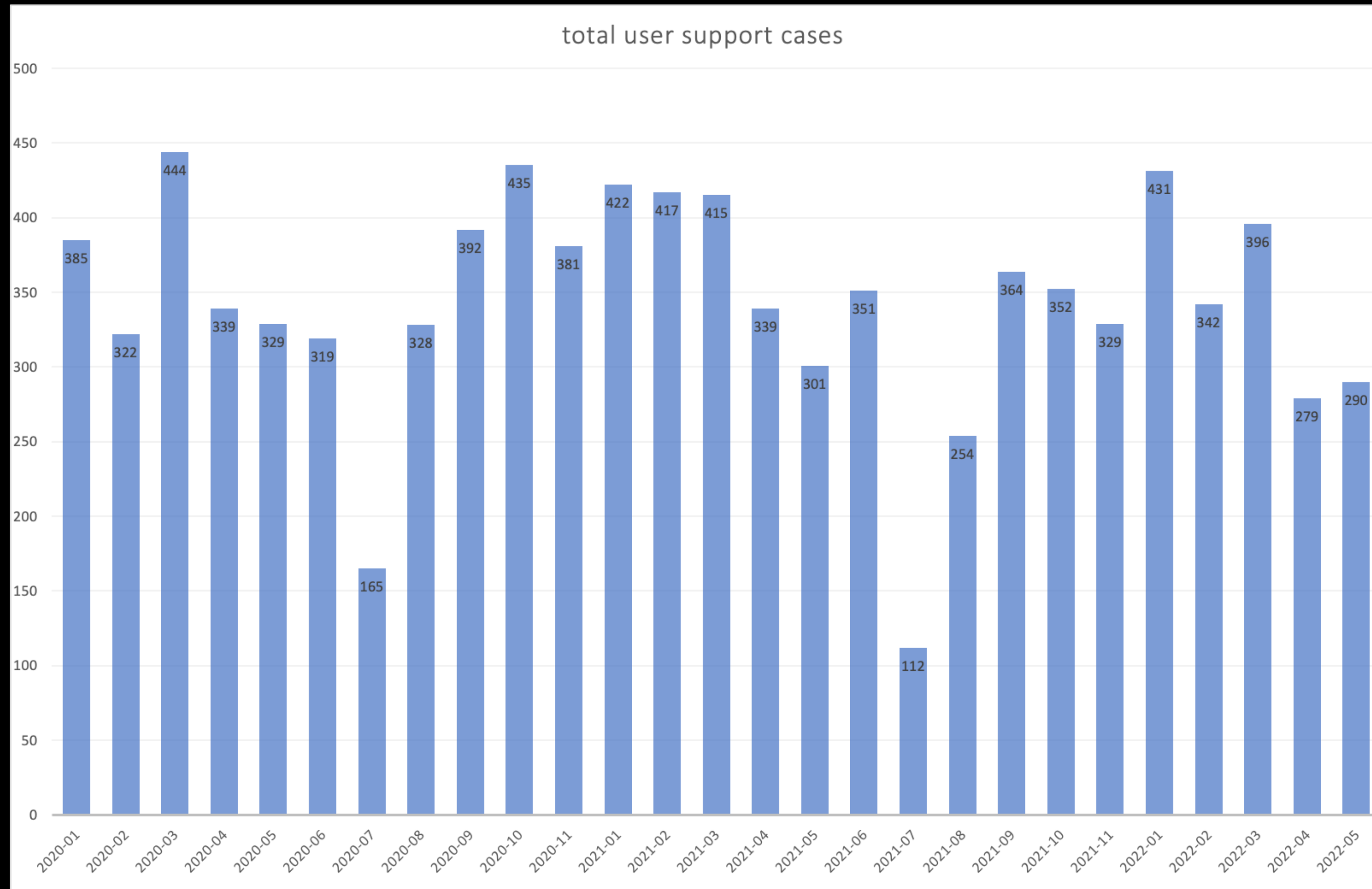- Re-use in Educloud

User Support

total user support cases

Growth
Projects - more than one a day

Growth
Users

User Support

total user support cases

Month	Cases
2020-01	385
2020-02	322
2020-03	444
2020-04	339
2020-05	329
2020-06	319
2020-07	165
2020-08	328
2020-09	392
2020-10	435
2020-11	381
2021-01	422
2021-02	417
2021-03	415
2021-04	339
2021-05	301
2021-06	351
2021-07	112
2021-08	254
2021-09	364
2021-10	352
2021-11	329
2022-01	431
2022-02	342
2022-03	396
2022-04	279
2022-05	290

Growth
Strategy for scalability and service quality?

- Listen to users, and stakeholders

- Secure financing (doubled UiO funding, project participation)

- Grow the team

- Collaborate and harmonise efforts

- Enable self-service

- Solve problems by addressing the root cause

Next
More about new features and services

- Self service

- Publication portal

- New Nettskjema integration

- Consent system

- Unified storage

Thanks, leoncd@uio.no

TSD and Educloud roadmap

2022 and beyond

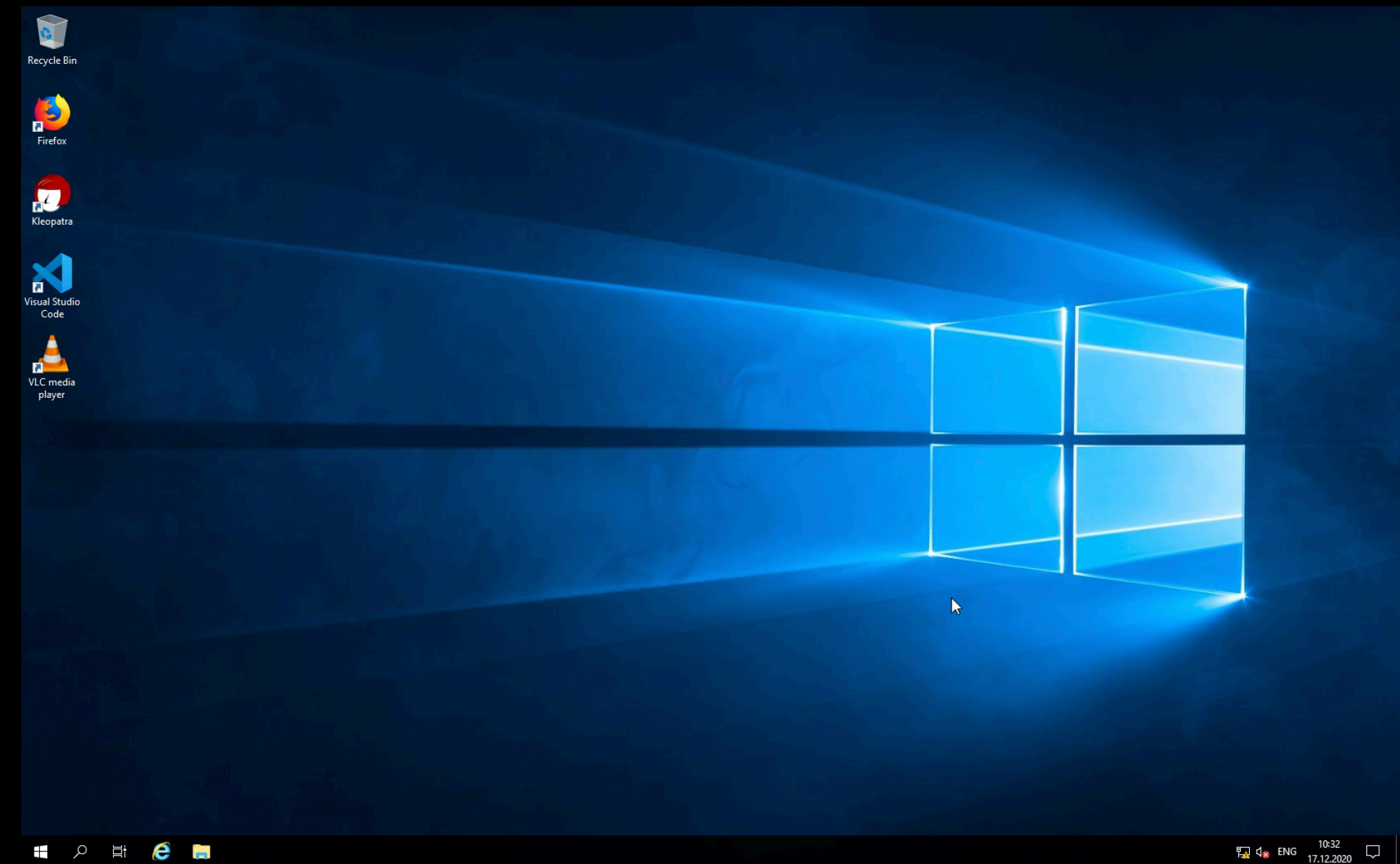**Leon du Toit - Section Leader, Research Services, USIT**

TSD and Educloud goals
High level

- Ensure security

- Manage growth

- Produce excellent and relevant research services

TSD

Enhanced self service

- Foreign users

- Resources

  - Storage

  - CPU

  - Memory

Storage

- Unified storage:

  - No more copying inside TSD to use HPC

  - Simpler operations

- New off-site backup

  - Disaster recovery

  - Tape storage

Export links

- The export equivalent of our highly successful import links

  - Create a link to a specific dataset

  - Set a password and and an expiry date

  - Share the link with a collaborator

  - Download data

Certification
IT Security and quality management

Virtual machines
More flexibility

- Try to re-use the "personal VMs" deployed in educloud

Team

- More new hires

  - HPC

  - Web development

Nettskjema

- Datasets automatically available, in near-realtime

- Self service for granular access

- Web portal (inside TSD) for accessing form data

- Software libraries

- Easier app development

- (Hopefully) integration with Helsenorge

Educloud

Feature parity with TSD

- Expanded self service (group management)

- Nettskjema

- Data portal (media capture, import and export links)

- Dynamic digital consent

- Publication portal

FAIR@uio

Thanks, leoncd@uio.no