THE NATIONAL IT INFRASTRUCTURE

Natalie Danezi

Big Data Processing in Astronomy & EO

23 January 2024

About SURF Infrastructure & Services

SURF IS A TOP LINE EXPERTISE AND RESEARCH CENTER

WE DEVELOP IT SERVICES FOR DUTCH EDUCATION AND RESEARCH

Service Provider

ICT infrastructure & services (ISO 27001)

Innovator

Push digital innovation & transformation

Association

Knowledge sharing: Expertise, training, meeting & support

Goals

Acceleration of member goals by collaboration

A note on the infrastructure

Data center: 100% renewable energy

GDPR data privacy
ISO 27001 information security

More than
200 Gbps High bandwidth network

More than
100.000 cores &
150PB storage

A note on the services

Storage & Data management

Compute

Security

Identity & Access management

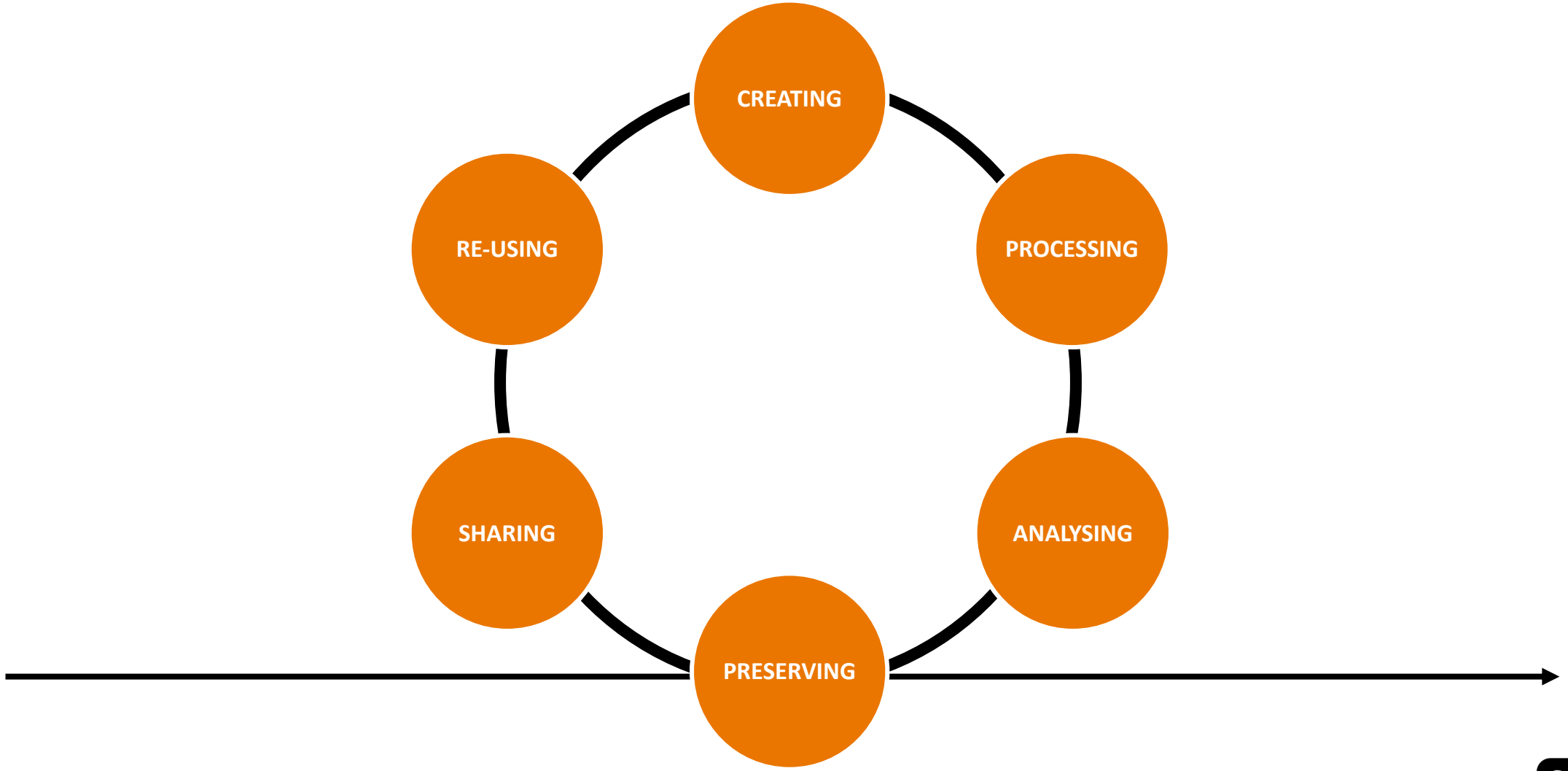
Network connectivity

Procurement & Contracting

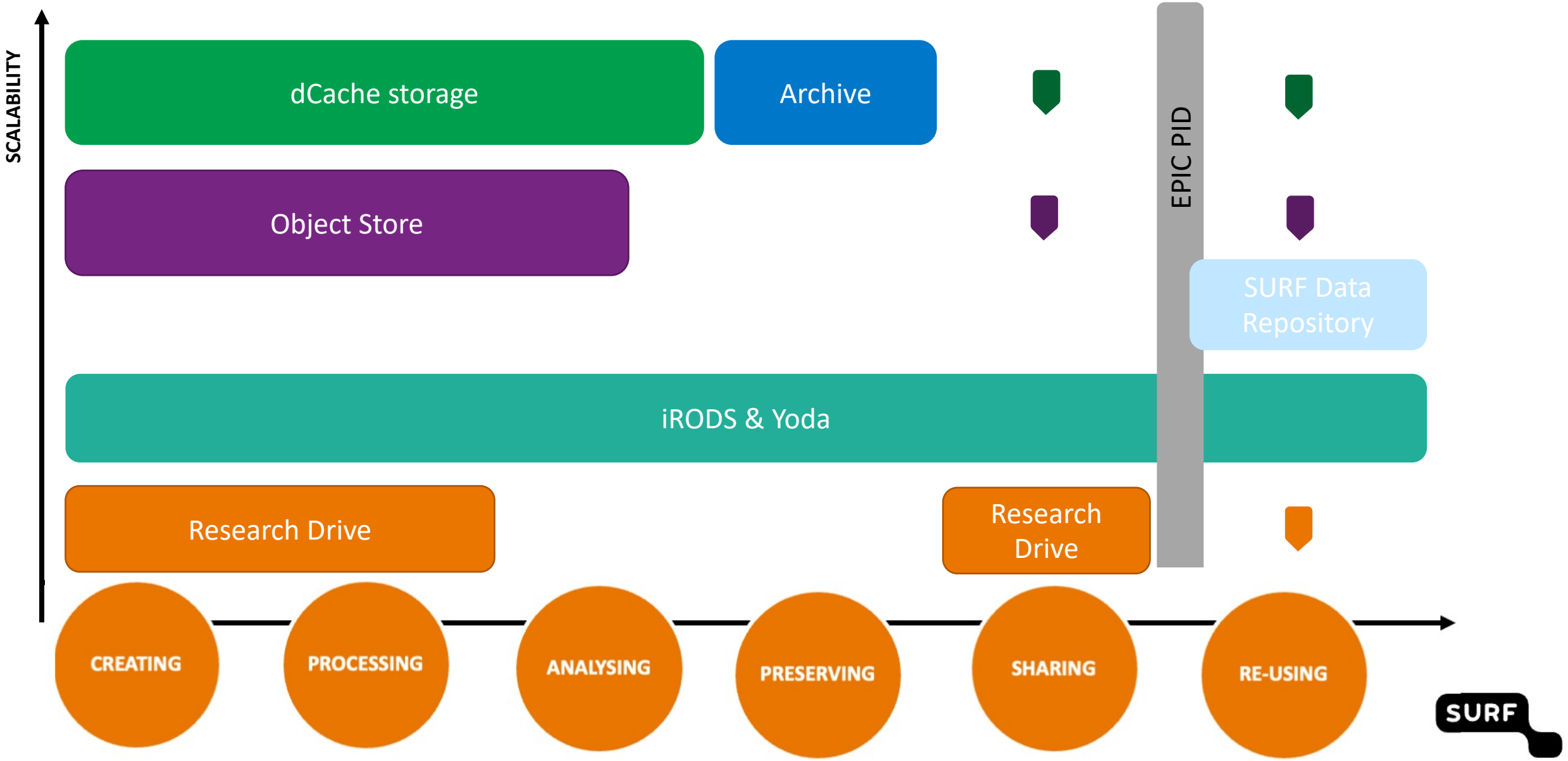
Storage & data management

New tape library 2022

Data Life Cycle – the holy grail of Research Data Management



SURF Data Services



COMPUTE

Snellius national supercomputer

High Performance Computing (Snellius)

Simulations and modelling that not only demand a lot of computing power and memory but also a lot in terms of communication between the various processors

Superfast:
Powerful HPC system (peak performance 14petaflop/s)

Machine Learning:
AI workloads on fast processors and GPUs with offered consultancy

Energy efficiency and sustainability are key drivers

Large collection of tools and libraries

Many cores (predominantly AMD)
Large symmetric multi-processing nodes
High memory nodes (4 TB and 8 TB)
Fast interconnects
A lot of work-space on disk
or a fast I/O subsystem
Phase 2 with higher CPU capacity

Data processing (Grid, Spider)

Data intensive projects processing instrument data from sensors, sequencers, telescopes, and satellites during the entire mission lifetime

Data volumes: parallel processing of large amounts of data, from many TB's to PB's

Processing pipelines: steady production workflows with semi-continuous data flows

Project organisation: international collaborations working on a shared set of data and sw

Ecosystem: modern cloud-based solutions with automated deployment

Built on the internal, elastic Cloud (OpenStack)
Fast network to external sources/dCache (1200 Gbit/s EVPN)
External connectivity for each compute node (2x25Gbit)
Fast local disks (up to 12 TB NVME SSD)
Powerful compute nodes (64 core, up to 1TB RAM)
GPUs: 16 A100 and 20 A10 GPU's

Spider Architecture

Interactive data processing
Interactive data visualization
Interactive monitoring

Interactivity

Event-driven processing
Customizable frameworks
Software portability & containers

Scientific workflows

Role-based project spaces
Data publication & redistribution
Access federation

Collaboration

User-facing utilities

Fast local disks for high I/O

Scalable Petabyte staging storage

Data Archive storage

External Scalable Distributed storage

Storage integration

Scalable batch processing cluster

Private project-tailored clusters

Private project-tailored nodes

ANSIBLE

Platform Deployment

Elastic Cloud

SURF

Terraform

Resource Provisioning

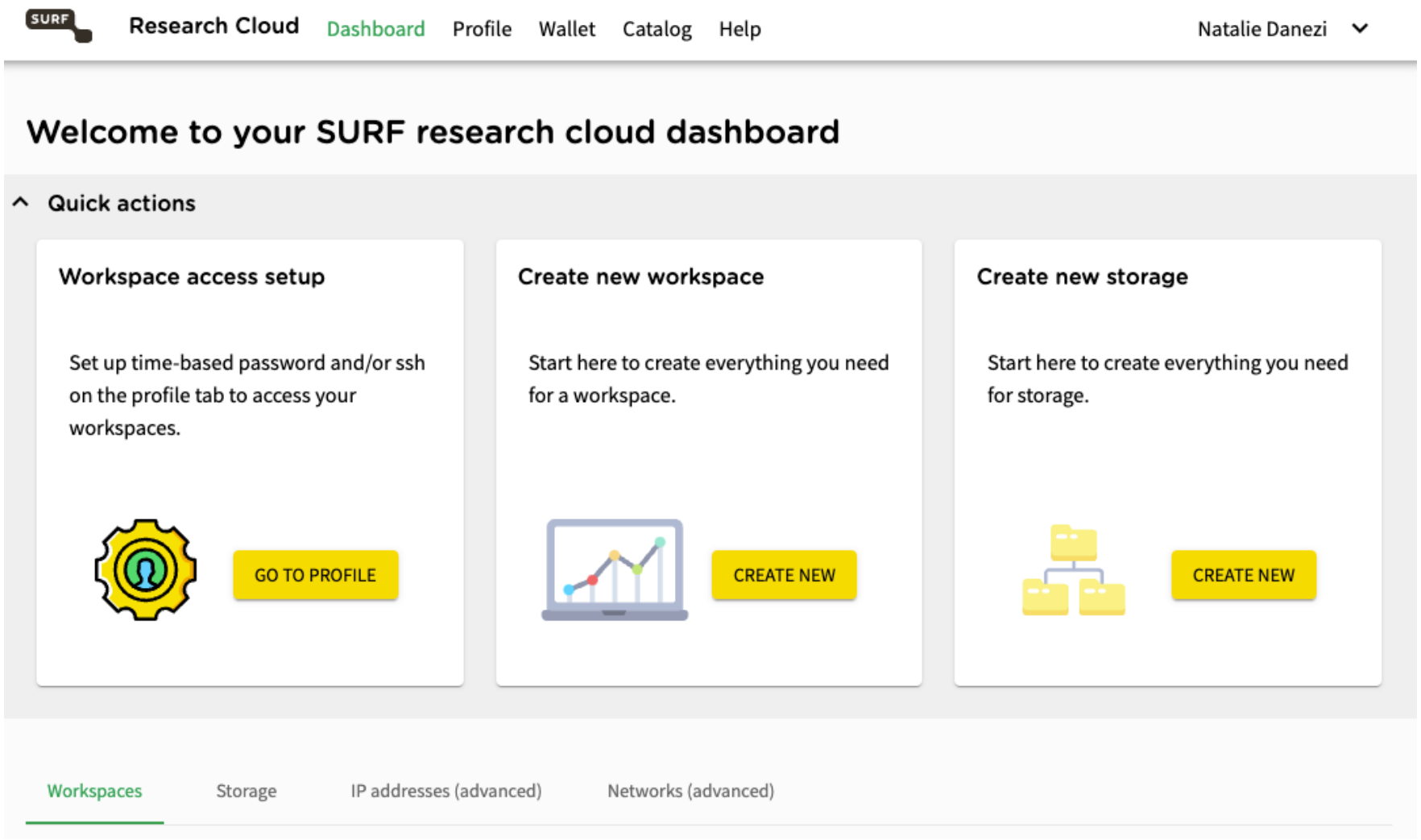
SURF

SURF Research Cloud (SRC)

Portal for building virtual research workspaces
(preconfigured or custom)

Virtualisation
On-demand
Flexibility
Self-service
Broker

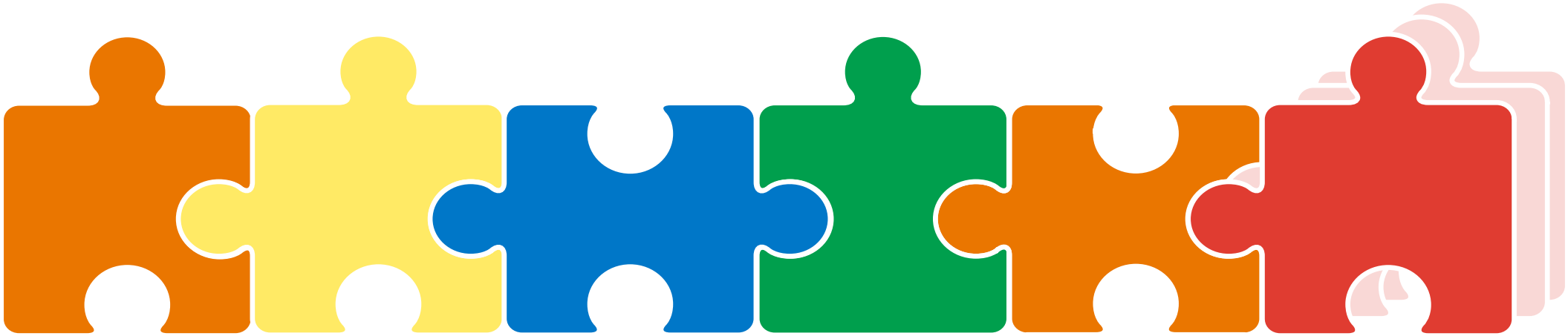
SRC environment



Cloud Research Consultancy (CRC)

- Tailor-made solutions
- Co-development
- Private, public or hybrid cloud infrastructure
- Long-term, sustainable production support

CRC Development Environment



Object store

Serverless
Functions
Containers

Queues

Streaming

APIs

Other managed components

Open call for scientific pilot project on the public cloud

edu.nl/4acht

Space Research Examples

SURF

WLCG/CERN


WEER | VERKEER | BEURS | VIDEO | FOTO | VKbanen

Volkskrant.nl

NIEUWS POLITIEK OPINIE BUITENLAND SPORT TECH & MEDIA

BINNENLAND | CULTUUR | ECONOMIE | REIZEN | WETENSCHAP & GEZONDHEID | OPMERKELIJK

'Tranen van geluk, Higgs-boson bestaat'

04/07/12, 12:03 – bron: ANP

© EPA. Applaus tijdens de persconferentie van het CERN.

UPDATE 'Wat een geweldige ontlading nu we weten dat Higgs-boson bestaat', zegt Jos Engelen vanuit Genève. Engelen is voorzitter van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en was voorheen directeur van het Europees Centrum voor Kernonderzoek (CERN) in Genève.

VERWANT NIEUWS

Higgs: nooit 04/07/12

Higgs of niet ontdekking

MEER OVER

Wetenschap

Ruimteteles ooit

100 jaar nad beroemde E geborgen

Vulkanoloog vulkaan en

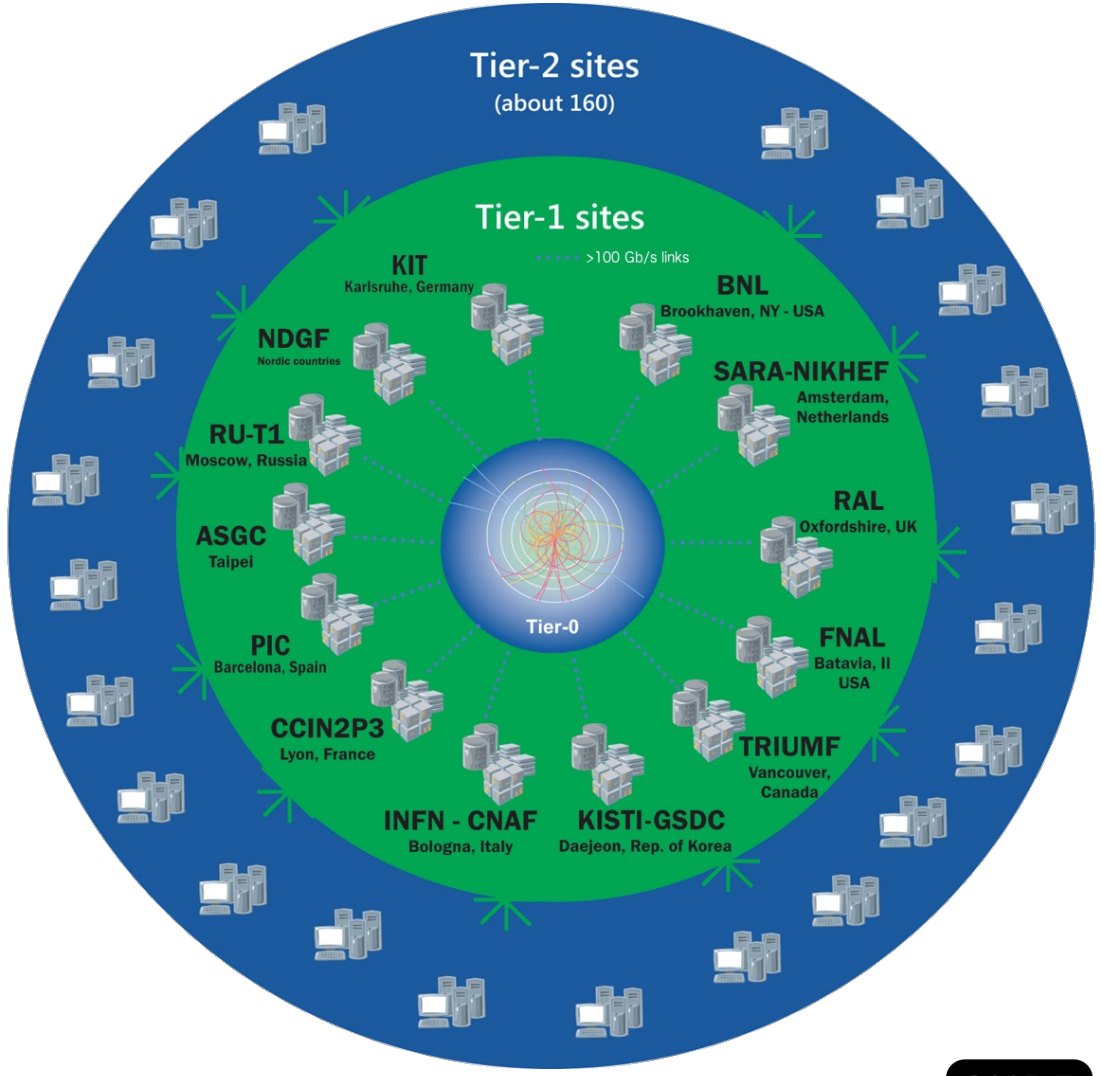
LHC experiments:
ALICE
ATLAS
CMS
LHCb

170 sites – Worldwide
42 countries
1.4M cores
2EB storage
12 000 physicists

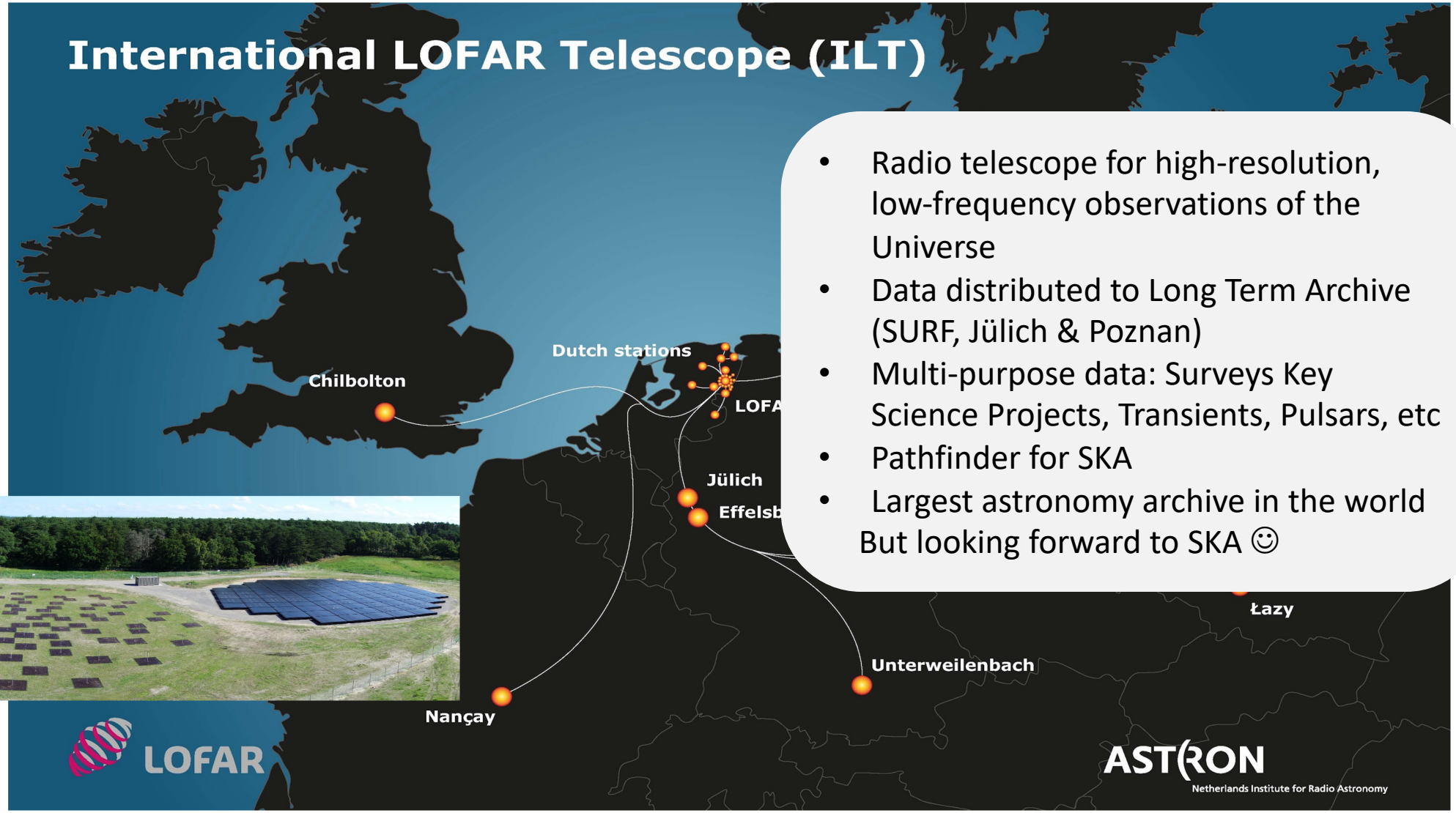
SURF contribution to CERN

Compute, storage, services
SURF and NIKHEF are a tier 1 site
Connected with dedicated, private, high-bandwidth network

Allocations 2023:
Core hours: 37M @SURF, 40M @Nikhef
Storage: 37PB tape + 13PBdisk @SURF, 3.5disk @Nikhef



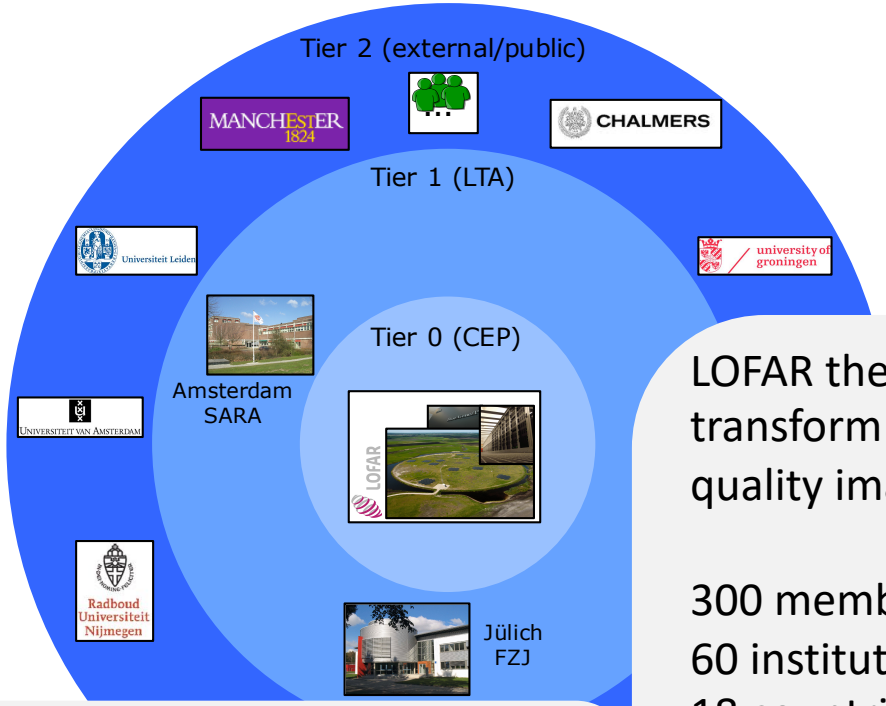
LOFAR

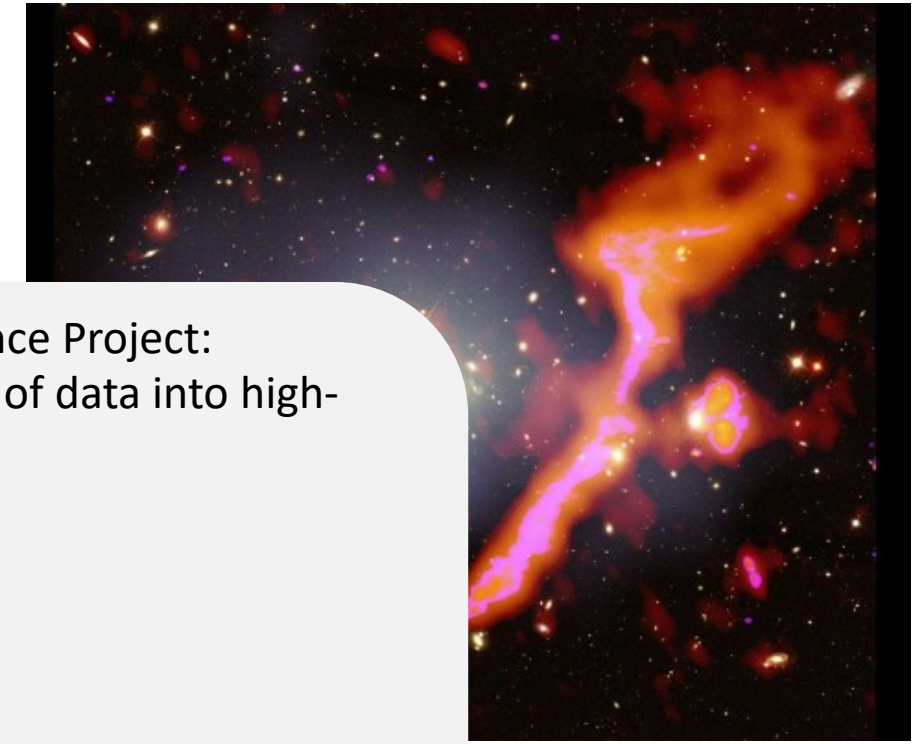


- Radio telescope for high-resolution, low-frequency observations of the Universe
- Data distributed to Long Term Archive (SURF, Jülich & Poznan)
- Multi-purpose data: Surveys Key Science Projects, Transients, Pulsars, etc
- Pathfinder for SKA
- Largest astronomy archive in the world

But looking forward to SKA ☺

SURF contribution to LOFAR





Compute, storage, services
30 PB stored at SURF
High-speed connection to the LTA
Processing power: Grid, Spider, Snellius

LOFAR the Surveys Key Science Project: transform massive amounts of data into high-quality images

300 members
60 institutes
18 countries
175 papers
Distribution of data products
Processed over 1000 LOFAR observations (4M core hours, 300TB disk and 700TB tape)
Science ready data products: publicly available through SURF data repository

LIGO-Virgo-KAGRA (LVK)

First GW signal from colliding neutron stars
Nobel Prize Physics 2017
Heavy computational load involved
10% of the processing on the Dutch Grid
Worldwide collaboration
Access to local & federated compute

SURF contribution to LVK

Compute, storage, services
Data retrieved from international cache servers
Hosting data distribution server for derived data

Allocation 2024:
Core hours: 24M @SURF and @Nikhef
Storage: 200TB disk @SURF and @Nikhef

The four gravitational wave detectors in the LVK scientific collaboration: LIGO at Hanford in the USA (top left), LIGO at Livingston in the USA (bottom right), Virgo in Italy (bottom left) and KAGRA in Japan (top right).

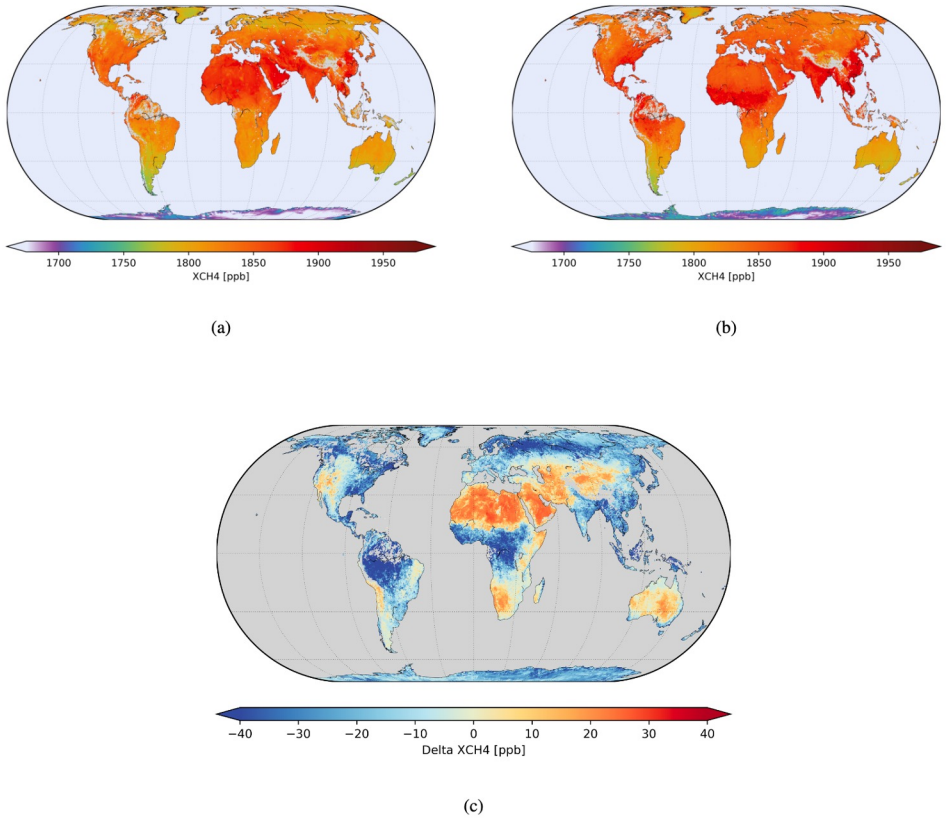
Tropomi

SRON and SURF collaborate on the scientific interpretation of the data by processing of incoming daily Sentinel-5P data and additional reprocessing

Detecting trace gases and aerosols in sub-city scale resolution

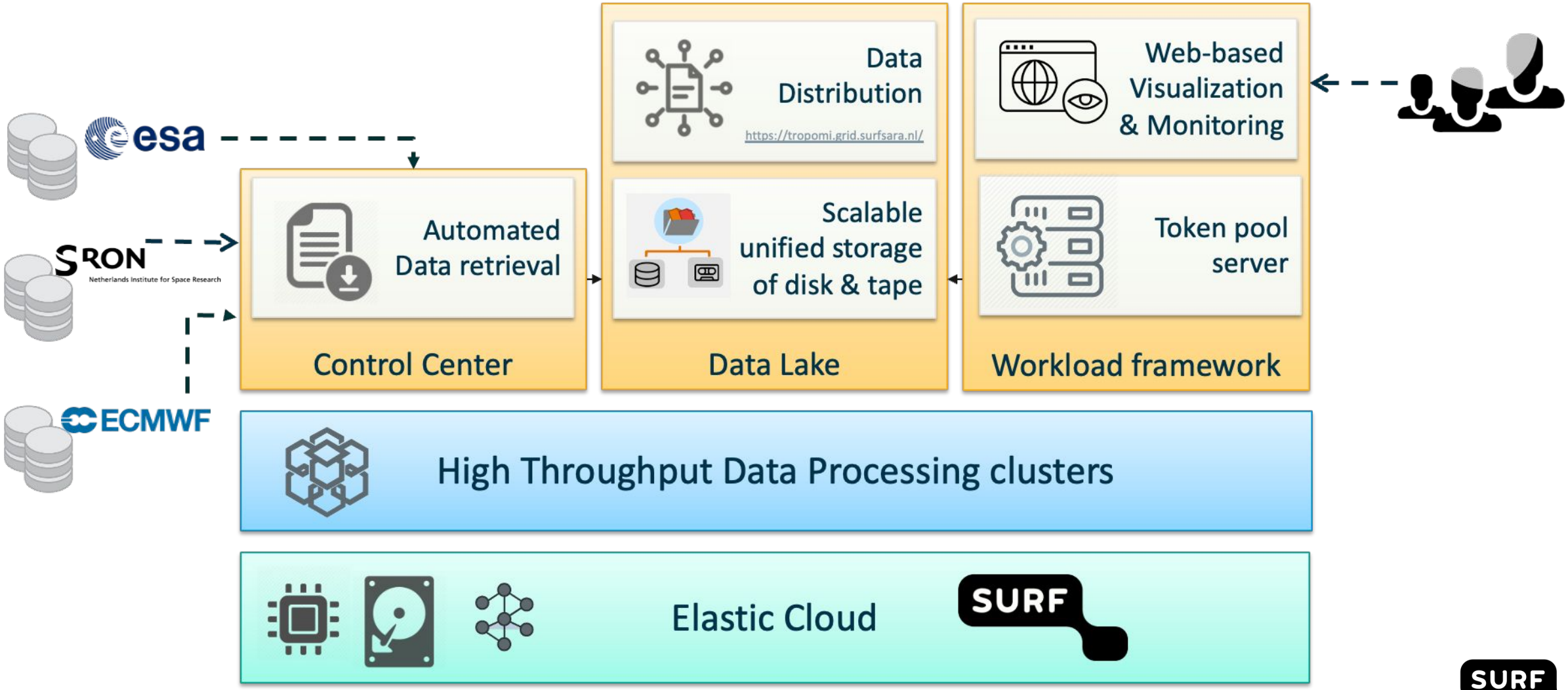
Extension of ESA products and extraction of as much scientific information as possible

Feedback to ESA proposing changes to the standardised data-processing algorithms



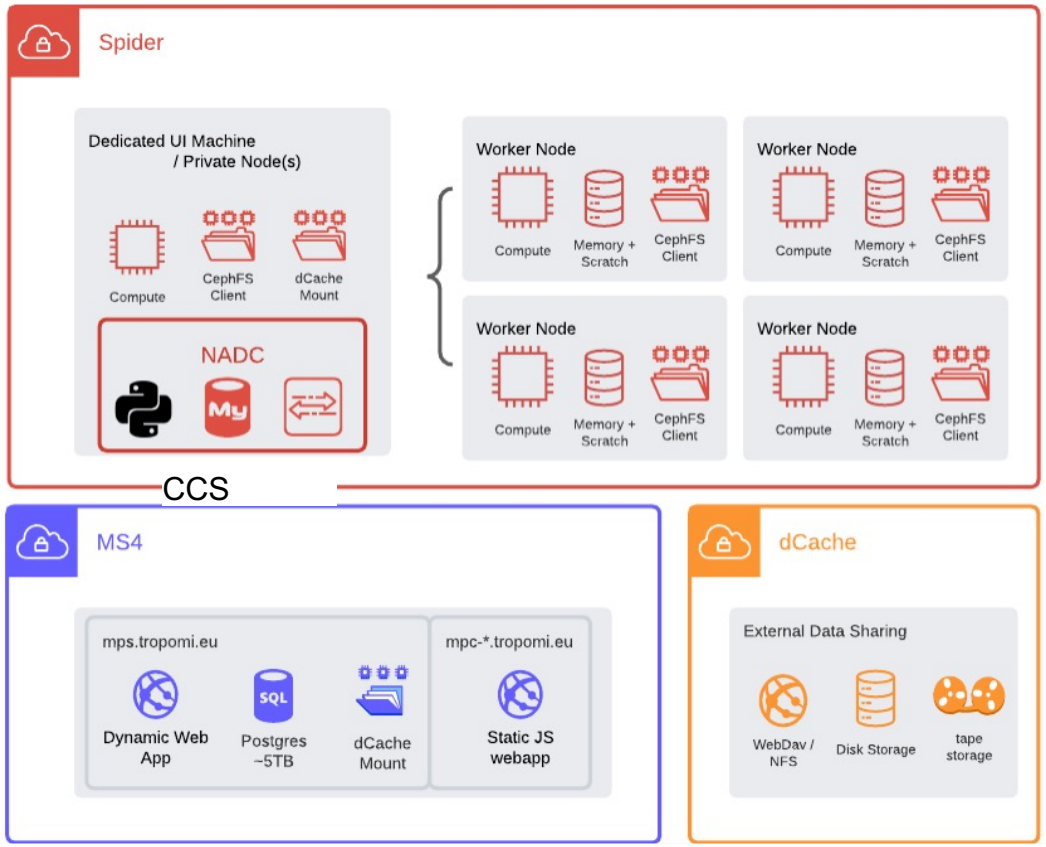
A. Lorente et al.: Methane retrieved from TROPOMI: improvement of the data product and validation of the first two years of measurements, 2020

SURF contribution to Tropomi



KNMI Tropomi Spider/CRC processing

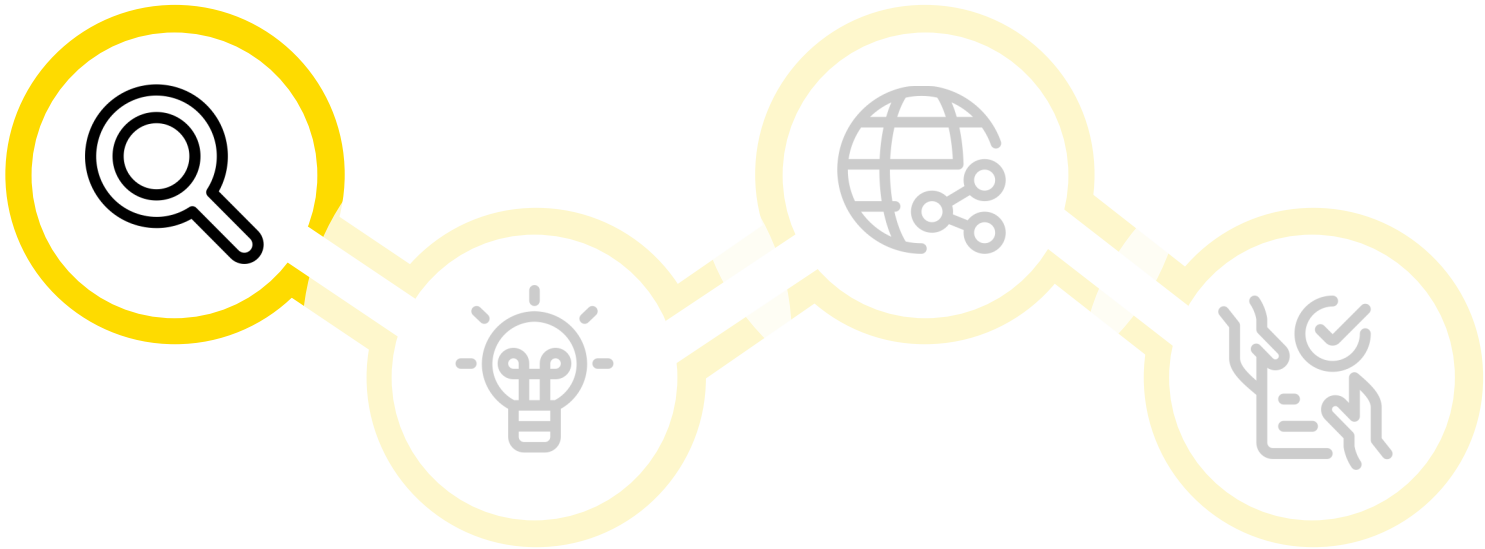
Custom platform
L1B/L2 satellite data processing
Integration of multiple services
Interoperable
Serving a large community
Fully automated batch processing
(data driven and time driven)



ROUTES TO SURF

1. NWO

Free at the point of use, including dedicated consultancy and storage



2. INNOVATION

SURF Labs – collaborations for the future

3. NATIONAL / EU / GLOBAL PROJECTS

SURF supports national (DCC, Roadmap), EU (C-SCALE) & global collaborations

4. CONTRACTS

There are (limited) possibilities to purchase (pre-)paid resources from SURF

National/EU/Global projects

EGI

GÉANT

Multiscale

C-SCALE

EUROPEAN OPEN SCIENCE CLOUD

EESSI
EUROPEAN ENVIRONMENT FOR SCIENTIFIC SOFTWARE INSTALLATIONS

CompBioMed

EUDAT

gaia-x

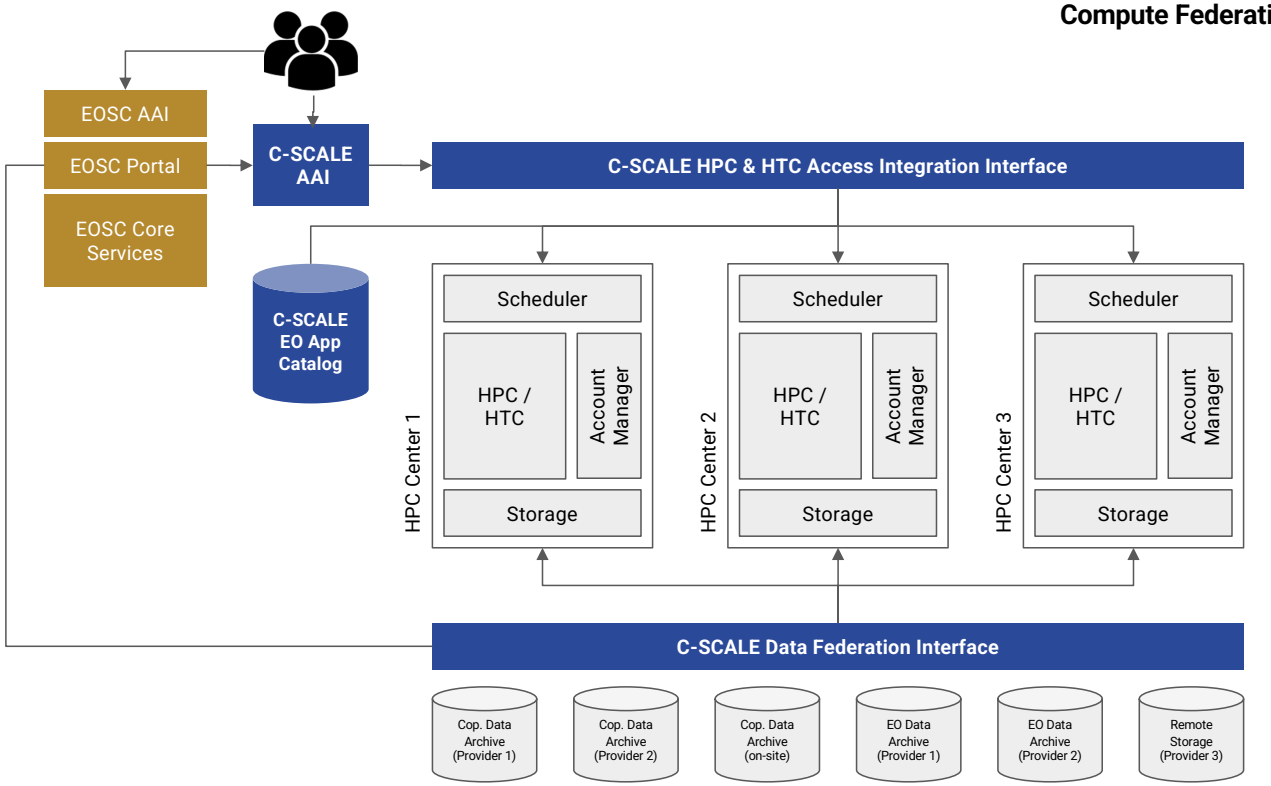
EURO

epi
European Processor Initiative

GREAT
Green Deal Data Space

SURF

C-SCALE & SURF: A new HPC & HTC federation



edu.nl/vqag3

Or contact:
raymond.oonk@surf.nl

Experimental Technologies Platform (ETP)

Explore the newest ICT technologies and methodologies that will shape the future of scientific applications. Experiment areas:

- **Benchmarking (FPGAs, graphics cards from AMD and Intel, smart NICs, data processing units)**
- **Workflow design and end-to-end optimization**
- **Architecture, provisioning, and orchestration**

edu.nl/kp8w8

Or contact:
servicedesk@surf.nl

Emerging technologies: processors, accelerators, storage and networking

Seminar

Seminar: Hardware innovation technologies for IT in research

In this seminar, SURF will present a number of hardware experiments that we are exploring with partners. These experiments aim to accelerate data-intensive research by exploring and testing developments in existing and new IT technologies.

Sign in

05 Mar 2024

11:00 - 17:00

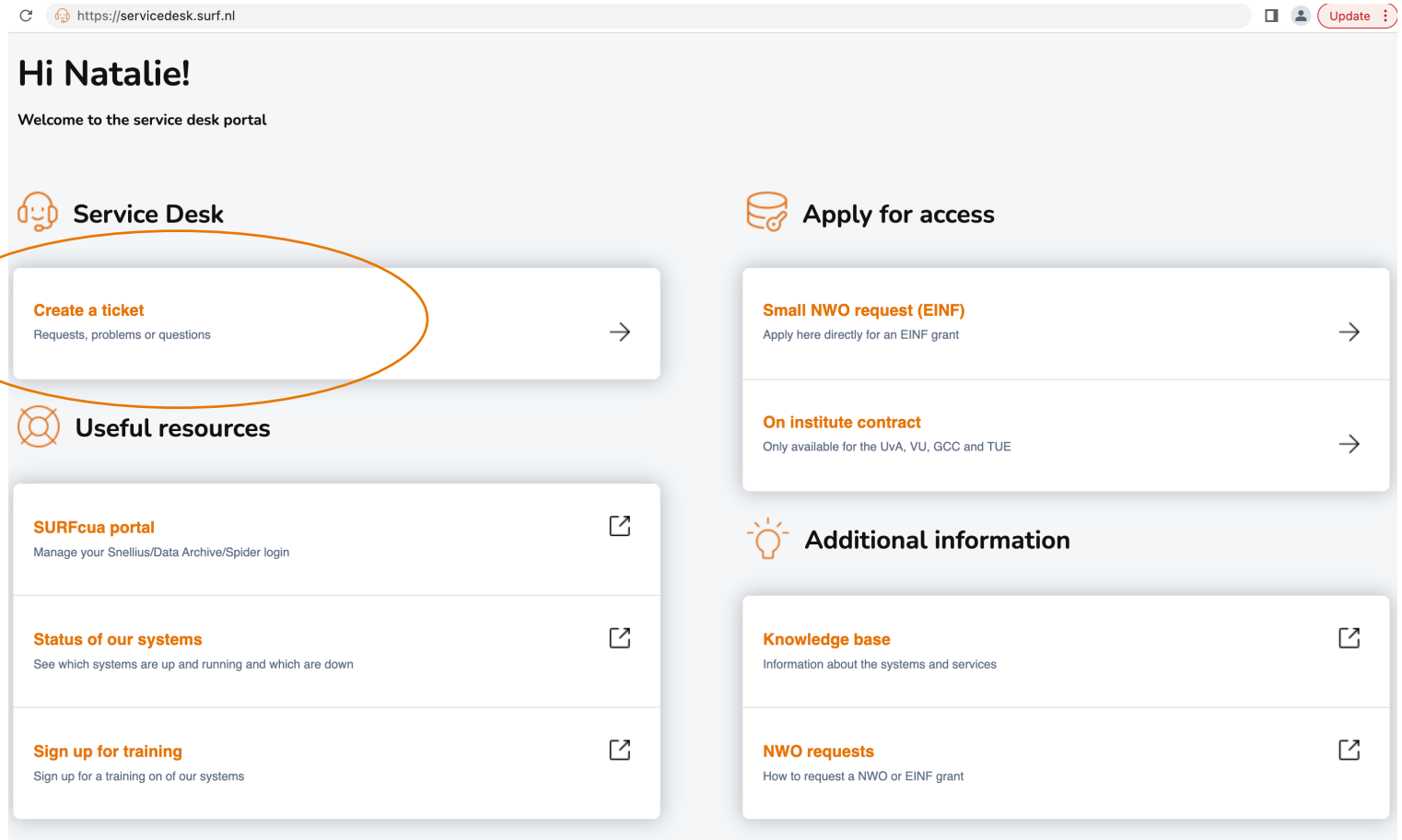
SURF Utrecht

Sign in

edu.nl/8r48d

Or contact:
raymond.oonk@surf.nl

Service Desk: for any question



edu.nl/h78ha

Or contact:
servicedesk@surf.nl

Thank you

: natalie.danezi@surf.nl

: surf.nl