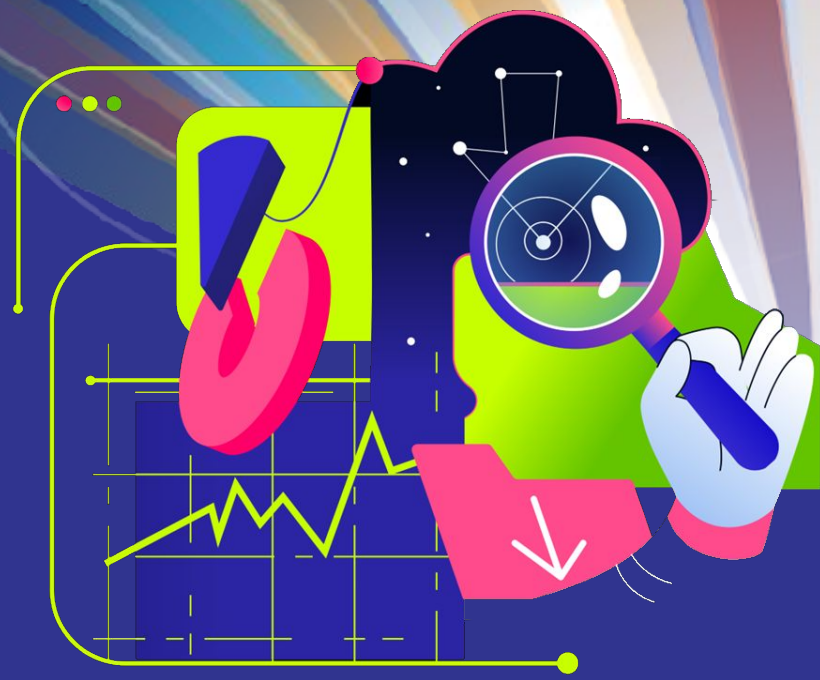
GÉANT – Data Autonomy

How to get more data autonomous at European Universities

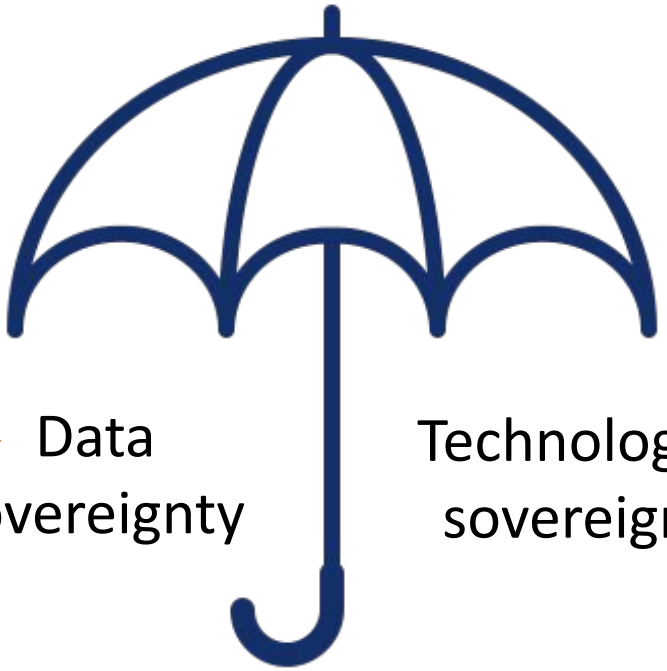
Erik Huizer
CEO - GÉANT

Launch Event Data Autonomy: Let's do it!
Groningen, 22 June 2023

Public



Digital sovereignty - a multi-dimensional concept



Retaining control or agency in the digital environment.

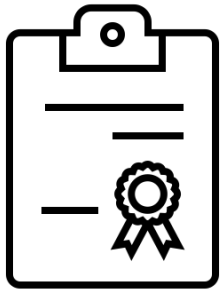
Control over the collection, access to, use and transfer of data as well as for subjecting data to the data governance structures of a certain community or jurisdiction

The power of states or communities to make decisions about how technologies are developed, used, accessed, distributed, and consumed

These concepts are sometimes subsets of each others, sometimes synonyms and are used differently depending on the context such as sector and the claims that are made about what it is to be sovereign.

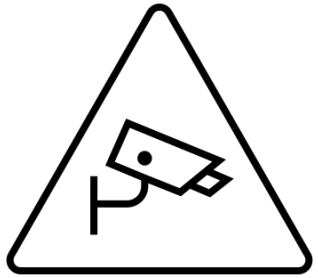
Based on: Digital Sovereignty, Data Governance and Researchers' Access to Data Part 1 (draft). (UvA – IvIR 2023)

Data Autonomy What is it and why?



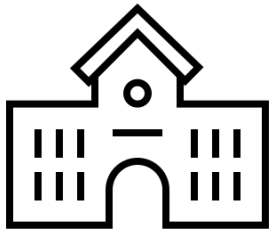
CONTROL

Understanding of ALL data assets



PRIVACY

Personal data (GDPR applies here)



POTENTIAL

SMART Campus / Research / Administration



SECURITY

Protection – threat surveillance/response



ACCESS

Identifying ownership and use



Data Autonomy What is it and why?



An Institute requires a carefully planned and implemented ***DATA ARCHITECTURE*** *in order to achieve Data Autonomy*

Data Architecture - Each and every data asset to be defined in terms of

Description (meta-data)

Data/data elements must be described in a consistent way that allows an understanding of the relationships between data assets.

Data policies and Access control

Data policies must describe how each data asset may be accessed and processed (and by whom).

Hosting

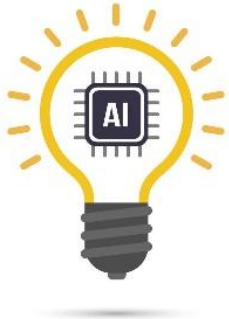
Data assets must be stored according to policy, be this on-premises; in the public cloud; in a community cloud. Data ownership and access policy does **not** vary across platforms.

Collection and curation

Data will be collected/produced in terms of workload (process support), but must be curated/modelled to reflect strategic relationships between data assets.

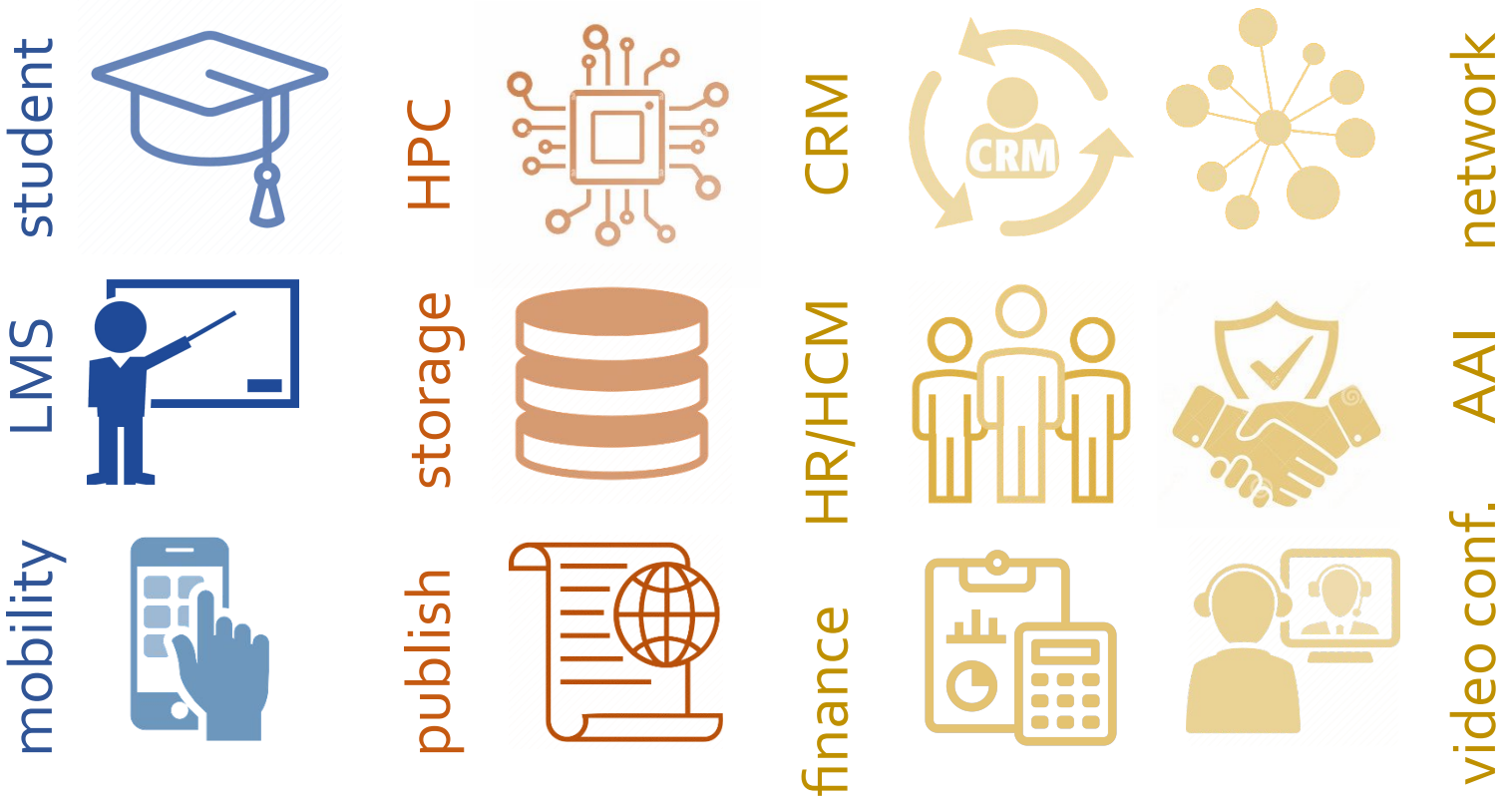
Potential

Data modelling will help identify opportunities for the institutional harnessing of anonymised; collectivised behavioural data in support of academic/research excellence.

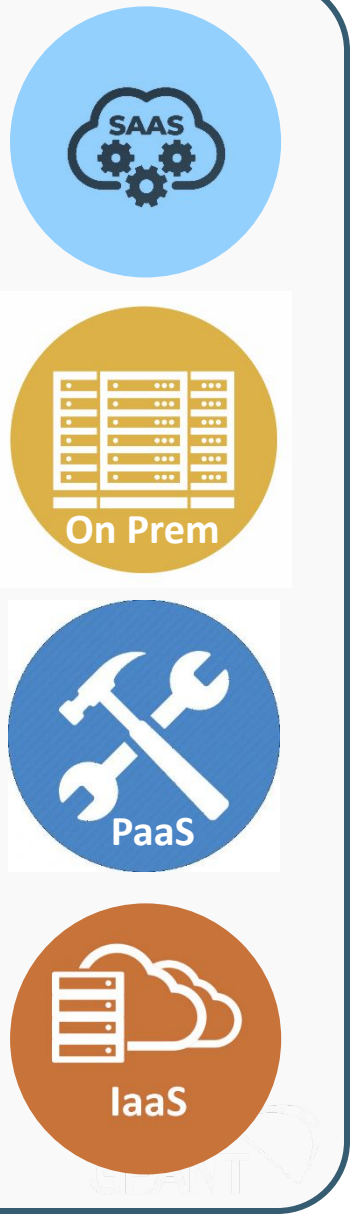


Data Assets - Curation and abstraction

Typical Uni workflows -



compliant
services
support



Data Assets – Hosting and processing

Enterprise workloads

- Most enterprise software is slowly moving towards consumable SaaS.
- It may be currently deployed in an on-premises D/C, or a private or community or public cloud (IaaS).
- Consumable SaaS provides APIs for the Unis to extract and curate strategic structured data.
- Legacy Business Intelligence solutions currently model data and provide strategic reporting.
- SaaS is by nature **GDPR compliant** and **secure** (surveillance on commercial networks).

Research workloads

- On premises hosting of HPC cluster (perhaps **hybrid**).
- Trending towards on-demand consumable **compute**.
- Research data storage in an on premises D/C, or private/community/public cloud (IaaS).
- NREN (eg SURF Research Cloud) provision of hybrid cloud resources (IaaS/PaaS/SaaS).
- PaaS providing bespoke research platforms (compute and storage) and containerisation.

*Cloud hosting does **not** grant data access to commercial companies*

Data Sharing – Open Science and Data Spaces

Data Autonomy facilitates collaboration via institutional **framework**

Data definition

✔ Data must be described using standardised meta-data in order to be **findable**

Access policy

✔ Is data Open; Restricted; Private. Restricted data accessed by request

Ownership

✔ Data ownership must be identified and data may be

Archival

✔ Data must be archived according to the definition and rules considered above

Storage

✔ Hybrid. Cold storage (meta-data online) cost effective and secure



European Data Spaces adopt the principles of 'Data Visitation' (ie. download unnecessary)

Zooming in on wishes and next steps in the Netherlands

Take aways from SURF's investigation on how to remain data-autonomous as a sector ("Regie op Data", 2022)

Frameworks and agreements based on guidelines, values and laws. This includes scoping and having a strategy related to looking after data.

Sharing and availability of data based on open and standardized metadata and architecture. Sharing across institutions on a national level.

Collaboration related to data sharing across institutional boundaries. Having rules and guidelines for international partnerships and incentives for sharing. Need for joint action.

Sovereignty and control: Clarity and agreement needed regarding who has control and responsibility over which data.

Infrastructure: Having a well connected (std's and architecture), well functioning (UI and functionality) federated infrastructure.

Awareness and digital proficiency: Needed for all (students, lecturers, scientists, policymakers, etc.)

Structural capacity: Having sustainable resources and funding.

Culture: We say we want to collaborate, but the actual willingness is low.

Time and money: Shortage of people dealing with these challenges. Temp-based projects.

The market: Big Tech is taking over and not necessarily supporting our values. Vendor lock-in and the affordability of the service is at stake.

Collaboration takes effort. Many different types of parties involved; who-does what and why?

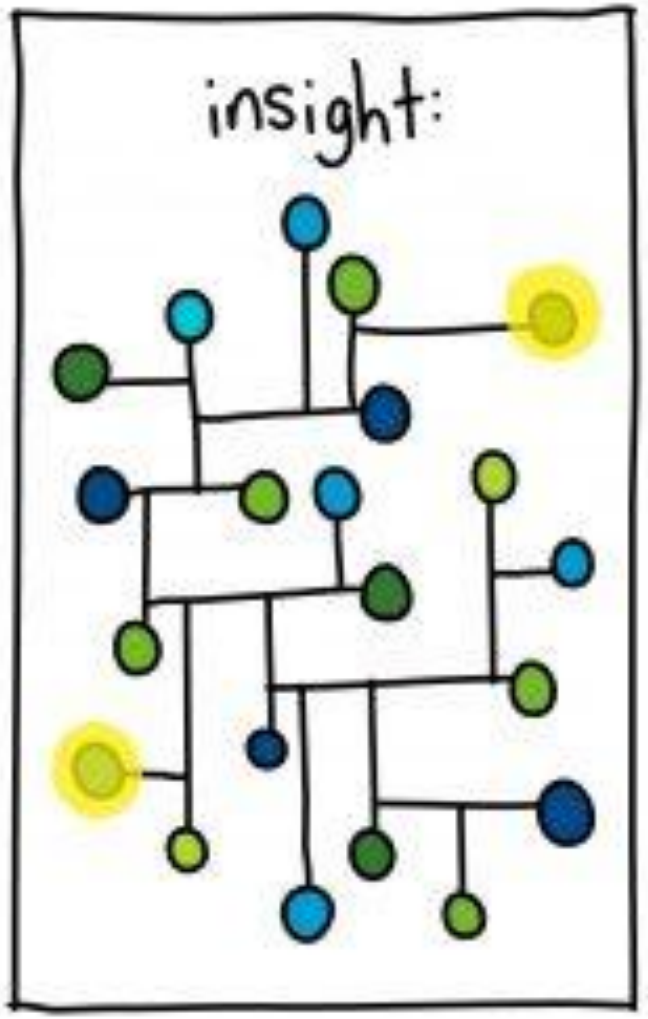
Data sharing sounds easy but is not. This is about types of data (research, study etc.), metadata, FAIR, as well as the culture of re-using data. Technical and mindset issue.

Knowledge and overview: The landscape of stakeholders and initiatives is fragmented and there is little overview (e.g. with decision makers). The world is changing fast, and we are not keeping pace…

Regie op data: https://www.surf.nl/en/5x-impact-in-2022-autonomous-sector-direction-with-members

Data autonomy initiatives to look into or follow:

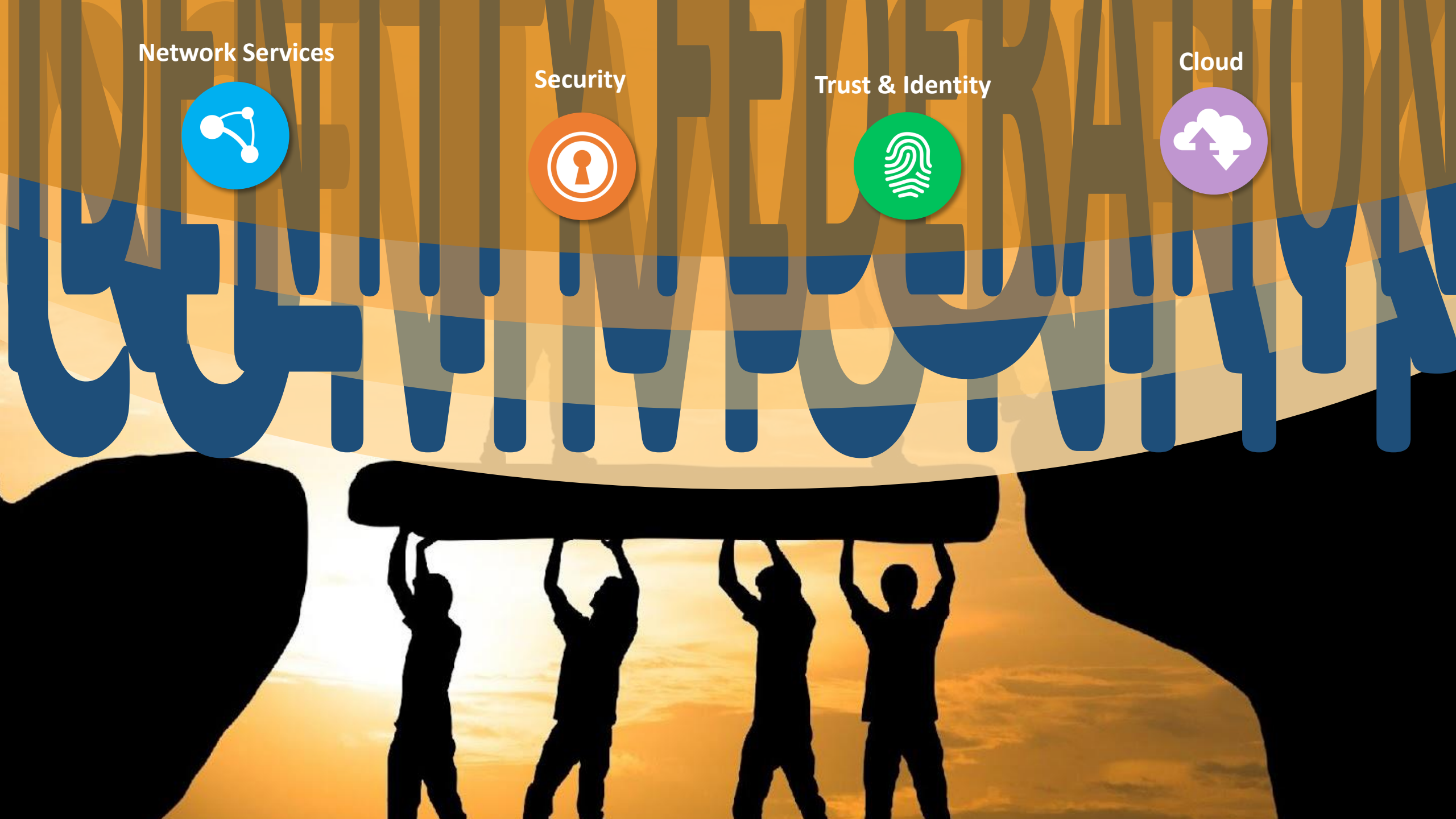
- **SURF Innovation zone Handling data responsibly**
  - Together with the member community, building a roadmap for the broader implementation of the FAIR principles, and supporting architecture and infrastructure.
- **LCRDM -** A working group of the National Coordination Point Research Data Management is addressing the question: how do we know we are in control of our research data? A practical guide to be published later this year.
- **Projects and activities** within The European University Association (EUA) and The League of European Research Universities (LERU)



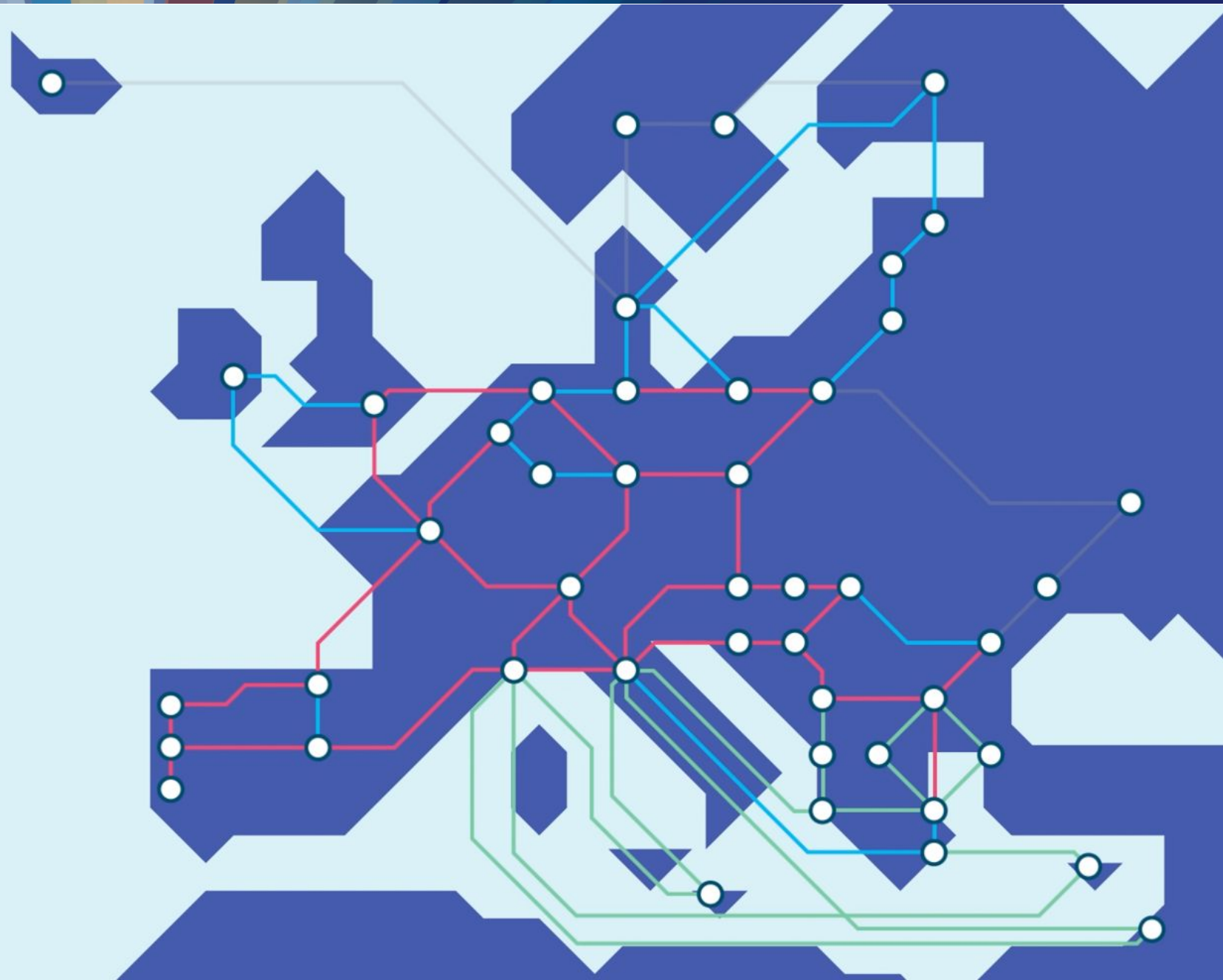
GÉANT Association supports and represents over 40 NRENs across Europe.

Together they support over 10,000 institutions and 50 million academic users.

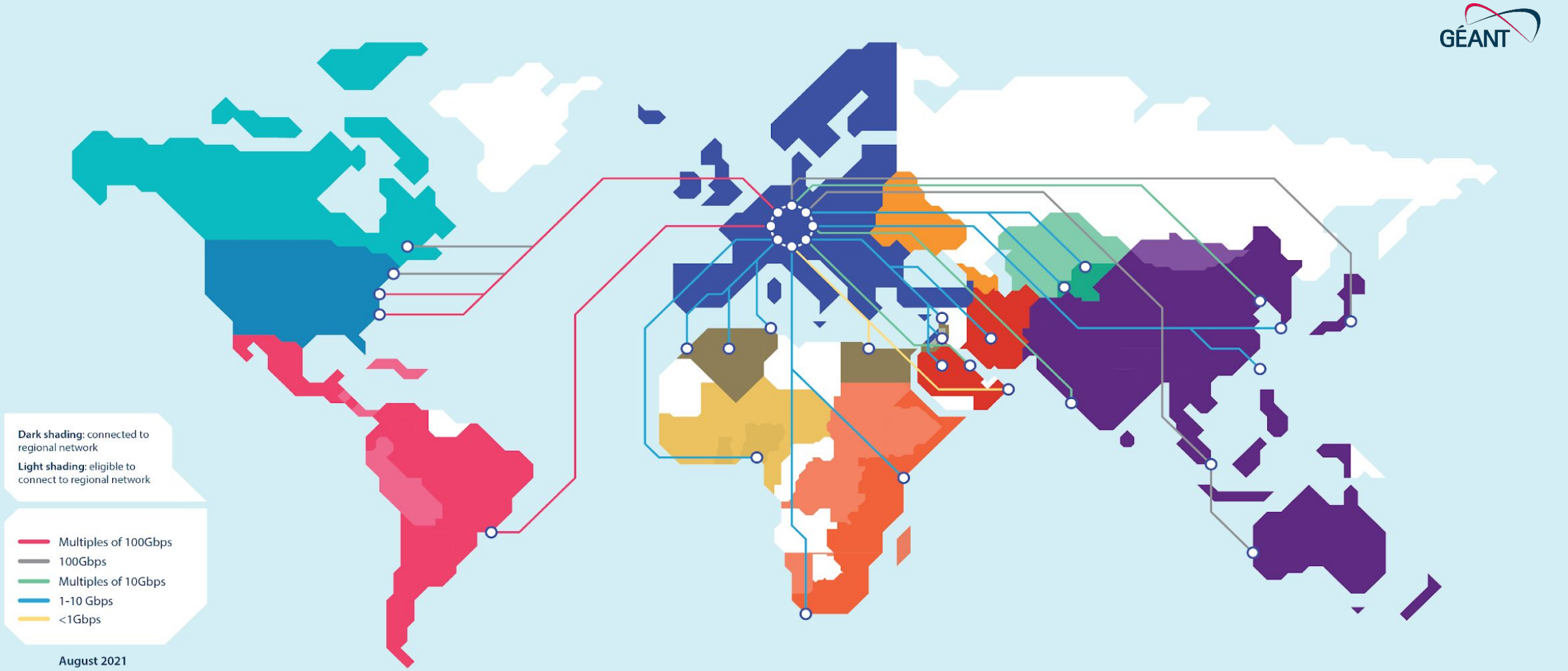



Network Services
Security
Trust & Identity
Cloud
INTEGRATION

GÉANT



Funded by the EU



Canada & USA
canarie
ESnet
INTERNET2

Latin America
Red CLARA

Europe
GÉANT
EU4Digital
EaPConnect

North Africa & Eastern Mediterranean
ASREN

West & Central Africa

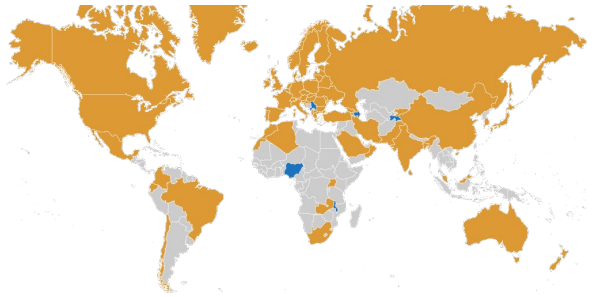
Eastern & Southern Africa
UbuntuNet Alliance

Central Asia
CAREN

Asia-Pacific
TEIN

Other R&E Networks

Federated Identities in Research and Education - Globally

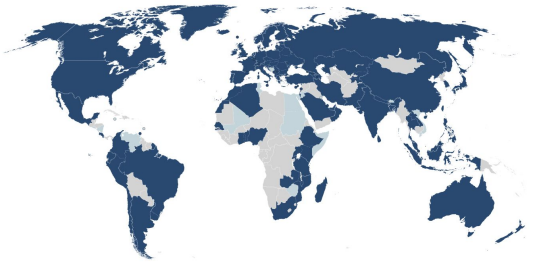


eduGAIN

80
Countries

5000+
Institutions

3500+
Web Services



eduroam

106+
Countries

9500+
Institutions

36500+
eduroam WiFi Locations

HPC

EuroHPC Joint Undertaking

FENIX RESEARCH INFRASTRUCTURE

LUMI

MyAccessID

Student Mobility

Erasmus+

erasmus without paper

MyAcademicID

EOSC

EUROfusion

umbrella

ExPaNDS

panosc photon and neutron open science cloud

SURF

LIFE SCIENCE RI

LS LOGIN

cessda

eosc AAI Federation

GÉANT

Core AAI Platform

eduGAIN

3d party (ORCID, eIDAS, Google etc.)

Astronomy/Physics

DOMA
Belle II
EUROPLANET
Rubin Observatory
SKA SQUARE KILOMETRE ARRAY
LHC
EHT
EUROfusion
iter china eu india japan korea russia usa
panosc photon and neutron open science cloud
CERN
ESCAPE European Science Cluster of Astronomy Particle physics ESFRI research Infrastructure
l'Observatoire de Paris
ESO
cta
WLCG

Earth Observation

EUMETSAT
Copernicus Europe's eyes on Earth
esa
ECMWF
Euro GEOSS
NOAA National Oceanic and Atmospheric Administration U.S. Department of Commerce
GEO GROUP ON EARTH OBSERVATIONS
NEXTGEOSS
ENVRI FAIR
EISCAT Scientific Association
cnes CENTRE NATIONAL D'ÉTUDES SPATIALES
EPOS ERIC European Research Infrastructure Consortium
WMO

eInfras

FENIX RESEARCH INFRASTRUCTURE
LUMI
PRACE
CloNeTS DS Clock Network Services Design Study
EuroHPC Joint Undertaking
QUANTUM FLAGSHIP

Life Science

EMBL-EBI
Roche
EOSC-Life
elixir
syngenta

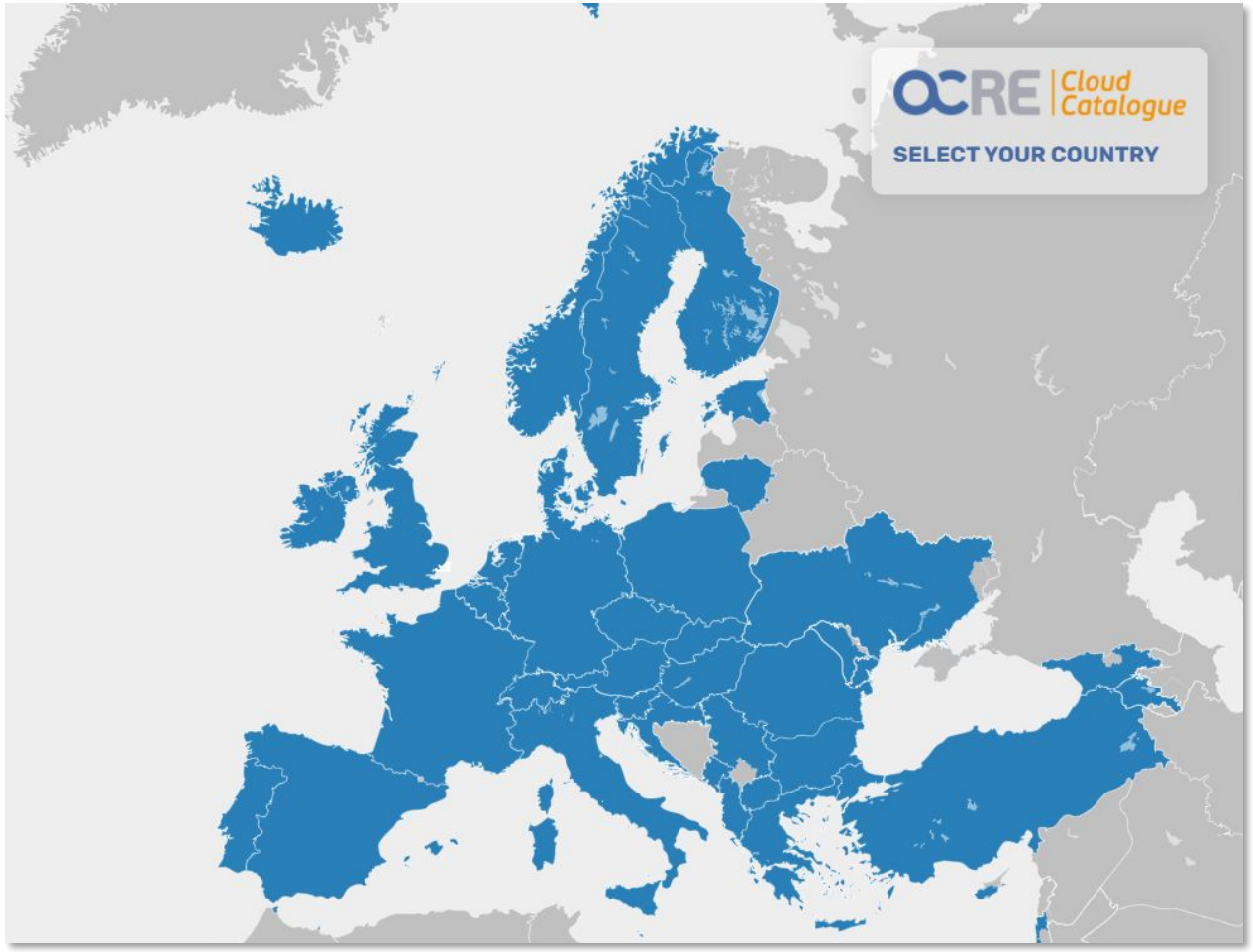
Arts and Humanities

SSHOC social sciences & humanities open cloud
USC Shoah Foundation
INPAPW

Outreach

GÉANT IMPACT
ENLIGHTEN YOUR RESEARCH
TF-RED Research Engagement Development

Procurement for Universities and Research Institutes in Europe



GÉANT brings down cost of cloud services

36 participating countries

R&E institutions act as a

Single Digital Market

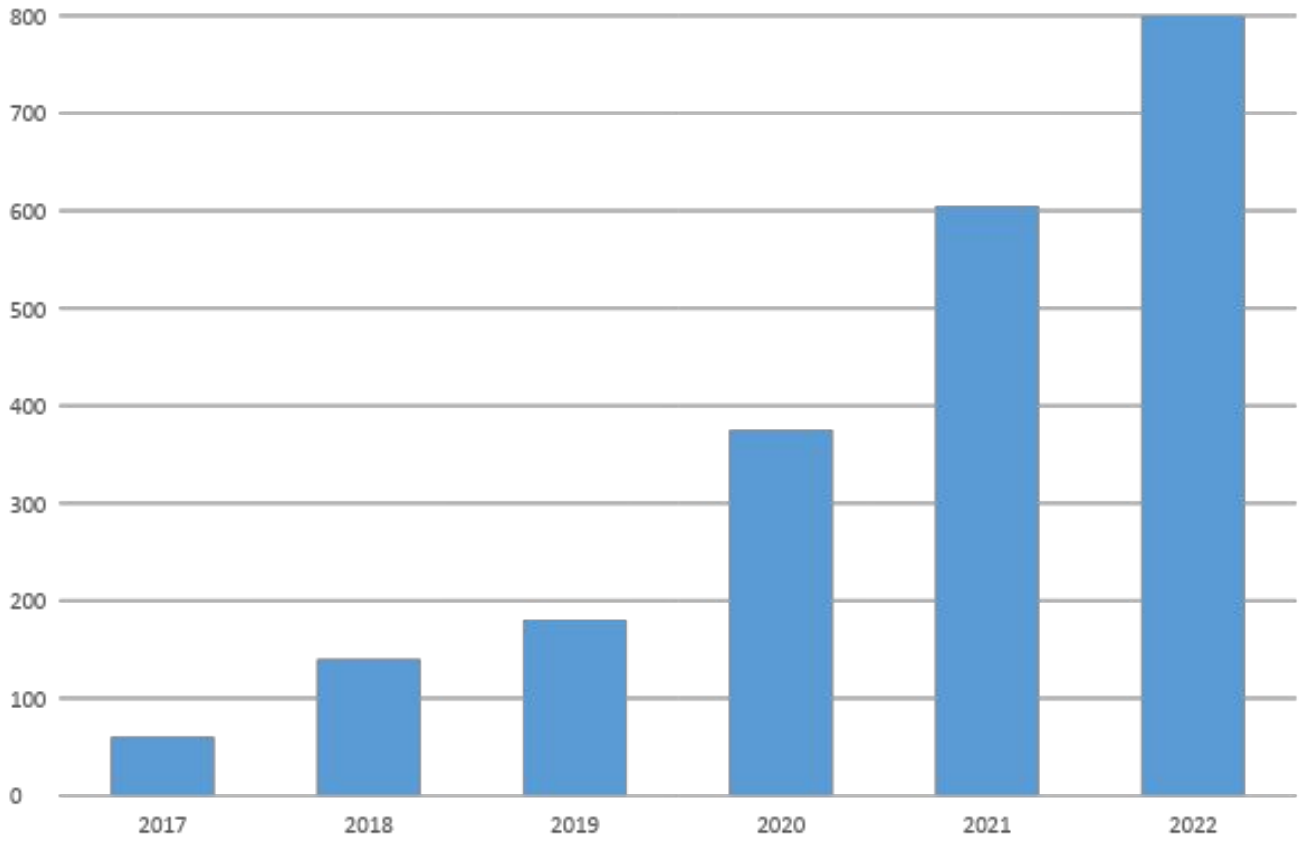
12 different IaaS solutions (providers)

20 suppliers,

incl. Amazon & Microsoft resellers

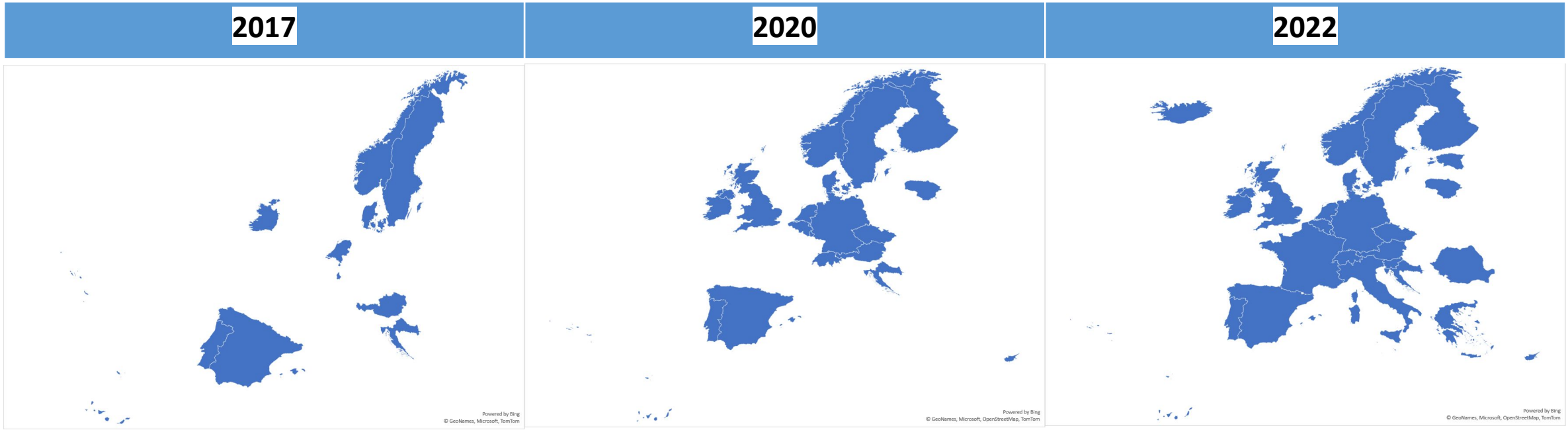
474 framework contracts issued

Cumulative growth in use of the GÉANT cloud portfolio



Number of institutions adopting the services has increased from 340 (2019) to 800 R&I institutions (2022)

Cumulative growth in use of the GÉANT cloud portfolio





Thank You

Twitter: @Milkshake

www.geant.org



© GÉANT Association
As part of the GÉANT 2020 Framework Partnership Agreement (FPA), the project receives funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 856726 (GN4-3).

InAcademia
ONLINE STUDENT VALIDATION

Service Uptake P3

9 federations

Country	Status
Denmark	Operational (May 2020)
France	Operational (March 2021)
Germany	Operational (Feb 2020)
Italy	Operational (June 2021)
Spain	Operational (Sept 2020)
Sweden	Operational (June 2021)
The Netherlands	Operational (May 2020)
Turkey NEW	Operational (November 2021)
Austria NEW	Operational (December 2021)

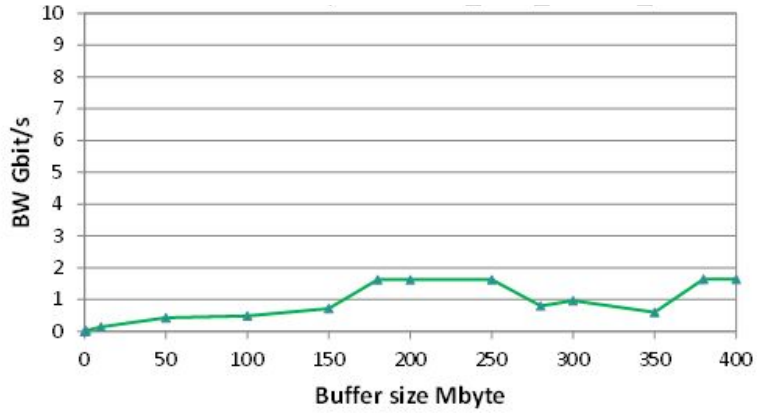


KPI	Baseline (2020)	Target	Measured
Number of student validations issued via InAcademia, measured annually	209,250	10% annual increase	2.3 million in 2022 526,669 Jan-Aug 2021

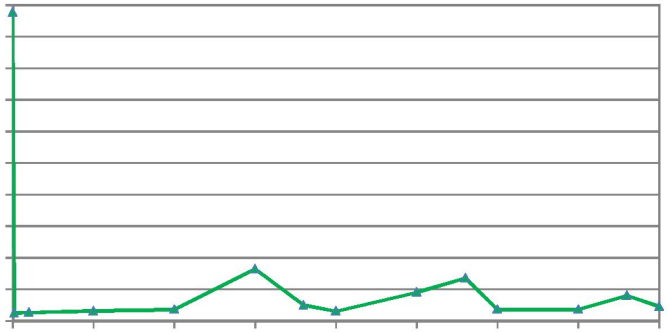
Optimised for research data transfers

Public Internet

Geneva to Canberra

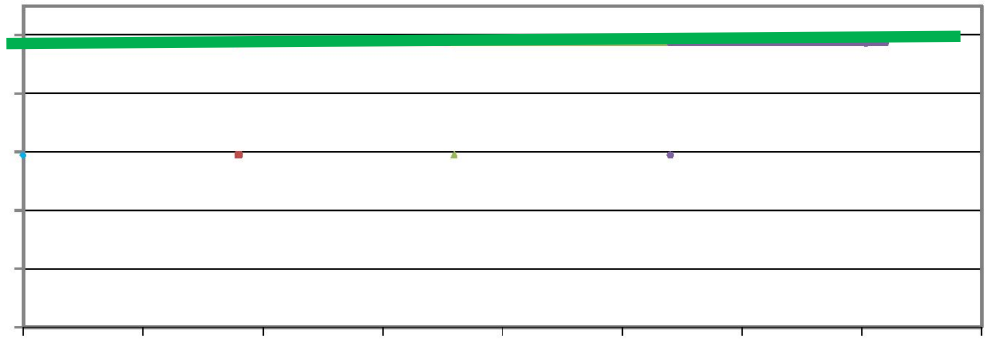


Geneva to Canberra



GÉANT and R&E partners

Geneva to Canberra GÉANT + R&E networks US to Australia



R&E networks are designed for different goals than the Internet

Comparative Times for a 100TByte data transfer.

File Size (TB)		Data rate (Gbit/s)	Time taken (Hours)	Time Taken (Days)
NREN	100	9.27	34.0	1.4
ISP A	100	1.72	183.2	7.6
ISP B	100	0.11	2864.3	119.3