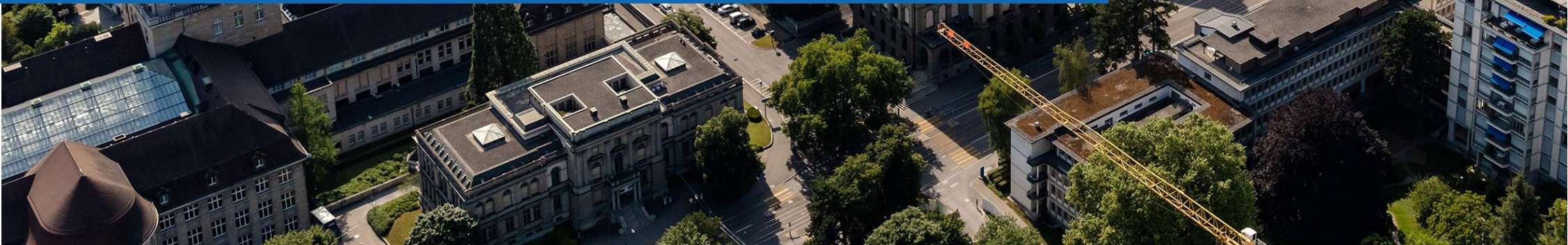


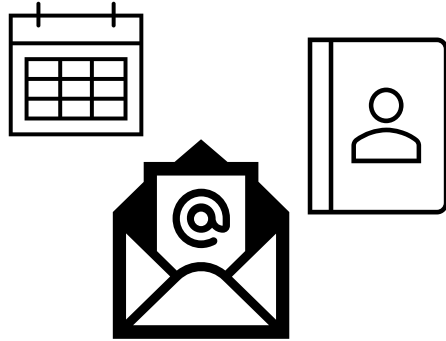


# Research Data in the Public Cloud

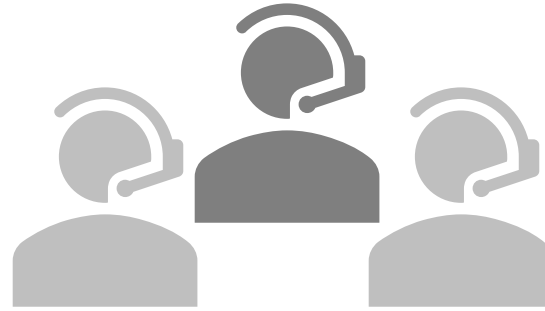
**Jarunan Panyasantisuk**  
Cloud Service Center, ID CCR  
EnhanceR Symposium, 7th November 2024



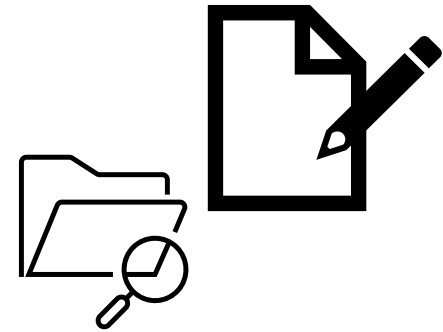
# Cloud services at workplace



Mailboxes,  
Calendars, Contacts



Communications



Office and  
collaboration tools

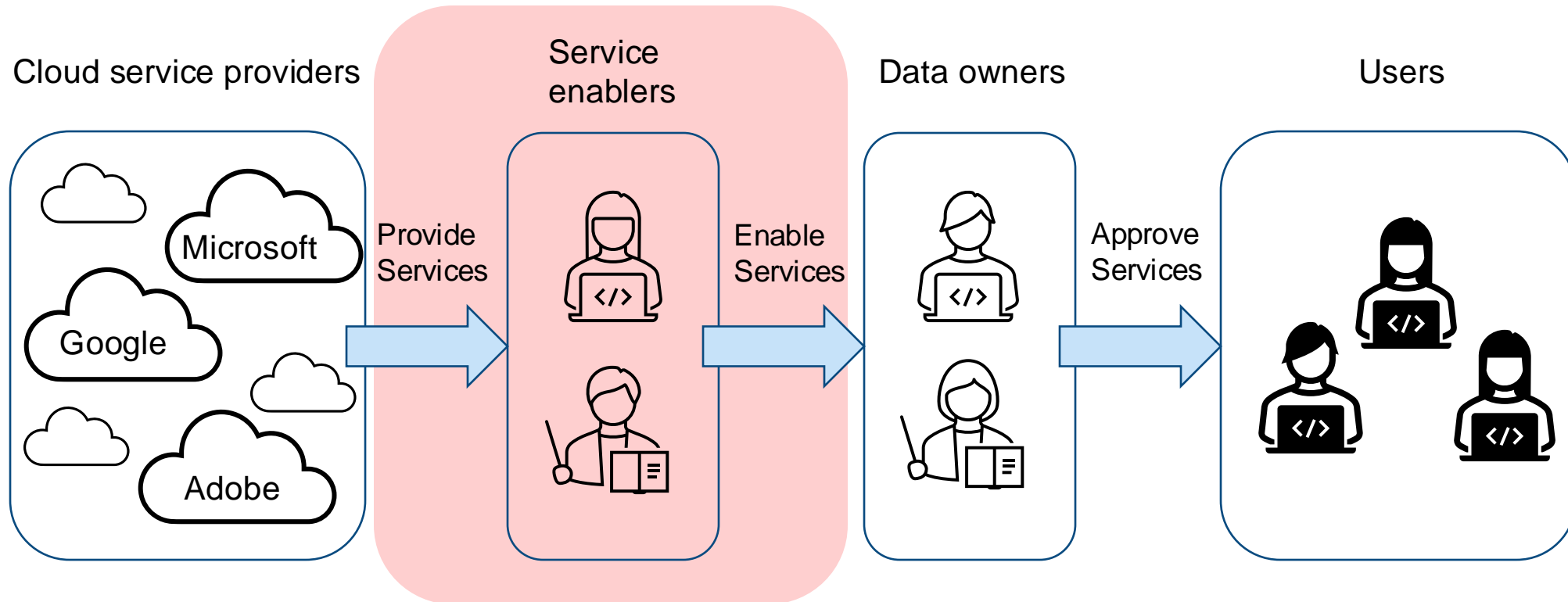
# About us

Research data in cloud infrastructure

Research data in cloud applications

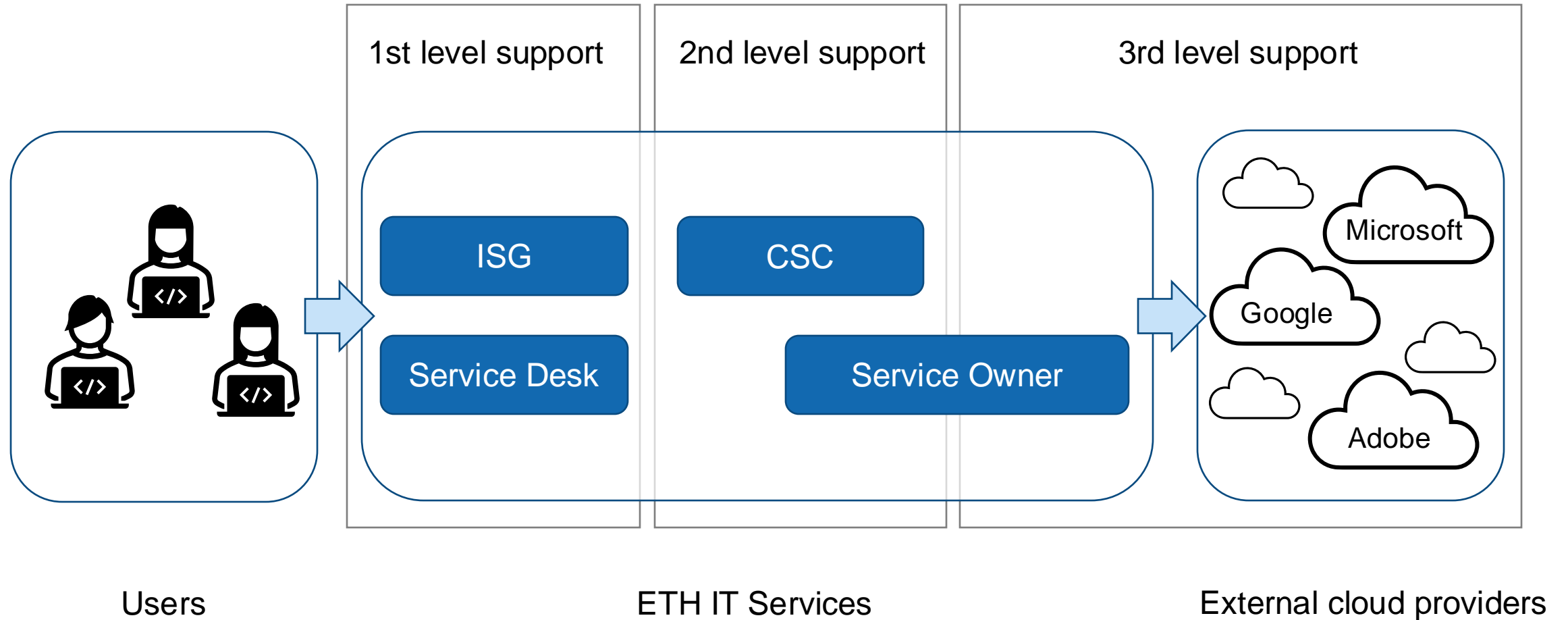
# ETH Cloud Service Center (CSC)

CSC enables and supports ETH members to use the cloud



<https://ethz.ch/staffnet/en/news-and-events/internal-news/archive/2021/10/secure-use-of-the-cloud/>  
<https://unlimited.ethz.ch/display/itwdb/Cloud+Produkte>

# Support system



About us

Research data in cloud infrastructure

Research data in cloud applications

# Data storage options

Local  
Computer

PolyBox

Gitlab

Local NAS

Central  
NAS, CDS,  
LTS

ETH Data  
Archive

Local  
Windows  
File Server

Cluster  
Storage

LibDrive

SaaS (M365,  
Google  
Workspace)

IaaS (Azure,  
GCP, AWS)

# When a user considers to use cloud infrastructure

- The user has use case in mind
- The user has to pay for the resource consumption

=> The user needs to estimate and compare the price

=> The user makes an intentional decision



# Data storage options

50 GB for each user,  
free of charge

PolyBox

Gitlab

<10GB free of charge

NAS: CHF125/TB/Year  
with Backup  
LTS: Free of charge

Central  
NAS, CDS,  
LTS

ETH Data  
Archive

Free of charge for  
published data

Euler: CHF200-400/TB/4y  
LeoMed: CHF200/TB/4y

Cluster  
Storage

LibDrive

Free of charge for  
published data

# Azure Storage

Popular
Compute
Networking
<b>Storage</b>
Web
Mobile
Containers
Databases
Analytics
AI + machine learning
Internet of Things
Integration
Identity
Security
Developer tools

## Storage Accounts

Durable, highly available, and massively scalable cloud storage

## Azure Data Box

Appliances and solutions for data transfer to Azure and edge compute

## Azure Data Share

A simple and safe service for sharing big data with external organizations

## Azure NetApp Files

Enterprise-grade Azure file shares, powered by NetApp

## Azure Stack Edge

An Azure managed device that brings the compute, storage, and intelligence of Azure to the edge

## Managed Disks

Persistent, secured disk storage for Azure virtual machines

## Azure confidential ledger

Tamperproof, unstructured data store hosted in trusted execution environments (TEEs) and backed by cryptographically verifiable evidence

## Azure Elastic SAN

Elastic SAN is a cloud-native storage area network (SAN) service built on Azure. Gain access to an end-to-end experience like your on-premises SAN

## Azure Managed Lustre

Azure Managed Lustre is a fully managed, cloud based parallel file system that enables customers to run their high performance computing (HPC) workloads in the cloud

## Azure Files

Simple, secure and serverless enterprise-grade cloud file shares

## Azure HPC Cache

File caching for high-performance computing (HPC)

## Azure Modeling and Simulation Workbench

Get Azure Modeling and Simulation Workbench pricing information. Onboard to the service with an Azure account and pay as you go for compute and storage use.

# Pricing comparison – Object storage

	<b>Azure Blob</b>	<b>Google Cloud Storage</b>	<b>AWS S3</b>
Frequently accessed (Hot)	259	279	331
Infrequently accessed (Cool)	168	190	219
Infrequently accessed (Cold)	113	101	-
Archive	55	28	55
Deep archive	-	-	24

- Costs in USD per 1TB per year
- Data centers in Switzerland

# Pricing comparison – File storage

	<b>Azure Files</b>	<b>Google Filestore</b>	<b>AWS EFS</b>
HDD Frequently accessed (Hot)	333	2,212 (min 1.1 TB)	5,160
HDD Infrequently accessed (Cool)	220		404
SSD	1,966	10,138 (min 2.7TB)	

- Costs in USD per 1TB per year
- Data centers in Switzerland

# Request examples

- A cold storage for 30TB
- Long-term use of GPUs for intensive computations
- Emergency short-term use of GPUs
- Blob storage for collaboration

About us

Research data in cloud infrastructure

Research data in cloud applications

# Using M365 and Google Workspace

## Pro:

- Convenient
- Easy to share objects with internal and external collaborators, e.g.,
  - Co-editing a document online
  - Sharing documents with a team
- Integrated with other online tools
- High availability
- Versioning
- Sync app with local computer
- Used to have large storage space

## What to keep in mind:

- Limited space
- Space extension with costs for group storage
- No central backup (yet)
- Not recoverable after retention period
- Data needs to be moved close to computing resources

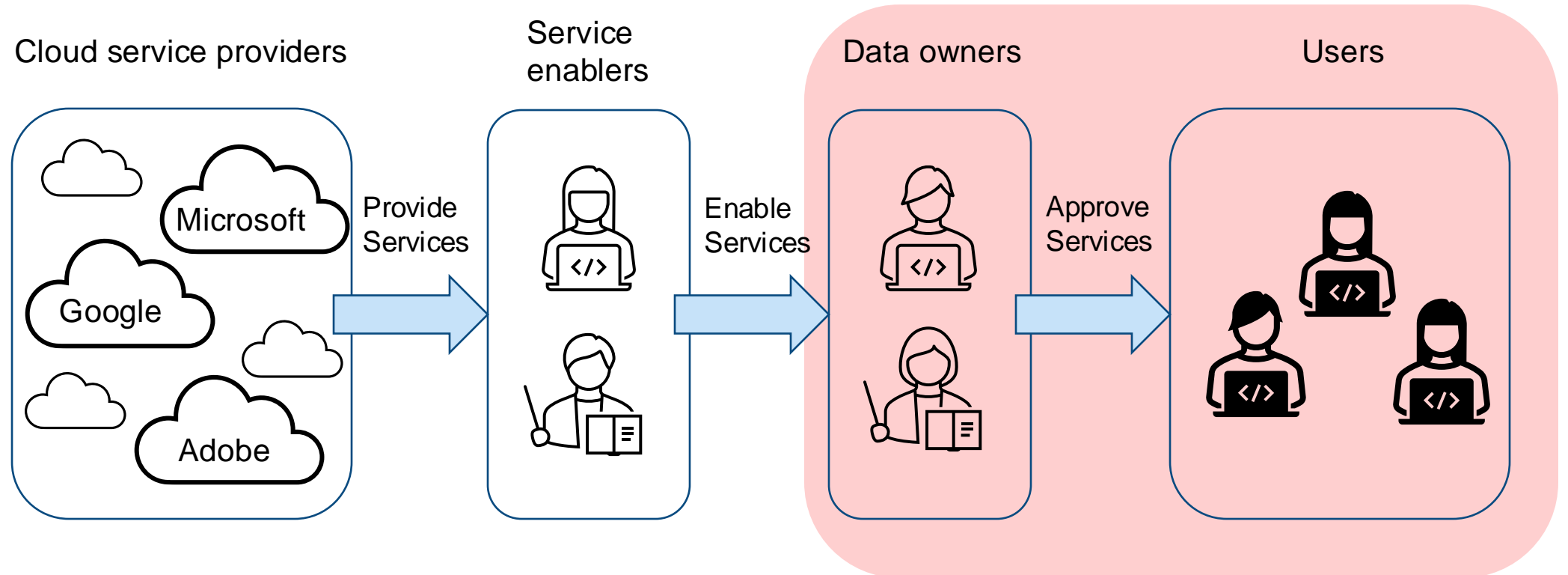
# Request examples

Requests	Status	Reasons
Using Zoom in MS Teams	Rejected	"Data storage happens outside ETH Zürich in data centers mainly in Europe (90%), <b>some in the US, and Australia</b> . Other locations might store data, too, depending on the residence of your host."
AquaMail	Rejected	"If you are located outside United States and choose to use the Service, please note that <b>your Personal Data are received and processed in the United States</b> , for which you have explicitly consented. "
Grammarly	Rejected	"Information submitted to Grammarly will be transferred to, processed, and stored in <b>the United States</b> ."



# Best practices

Know your data and how it is being handled in the cloud



<https://ethz.ch/staffnet/en/news-and-events/internal-news/archive/2021/10/secure-use-of-the-cloud/>  
<https://unlimited.ethz.ch/display/itwdb/Cloud+Produkte>

# What happens with the data in the cloud?

## Information security:

- Is my data classified as internal, confidential or strictly confidential?
- Which data is allowed to be stored in this cloud service?

- Where is the data location?
- Which data is collected and processed?
- Would the data be shared to third-party companies?



## Storage purpose:

- How large is the storage quota?
- What type of data should be stored in this cloud service?
- How much does it cost?

## Data lifecycle:

- What happens when my data is deleted?
- And when my account is deleted?

**Thank you for your attention!**

**jarunan@ethz.ch**