



Secure Handling of Confidential Research Data

SRDD2020

Dr. Diana Coman Schmid

Service Manager, Personalized Health Data Services

Scientific IT Services, ETH Zurich



Agenda

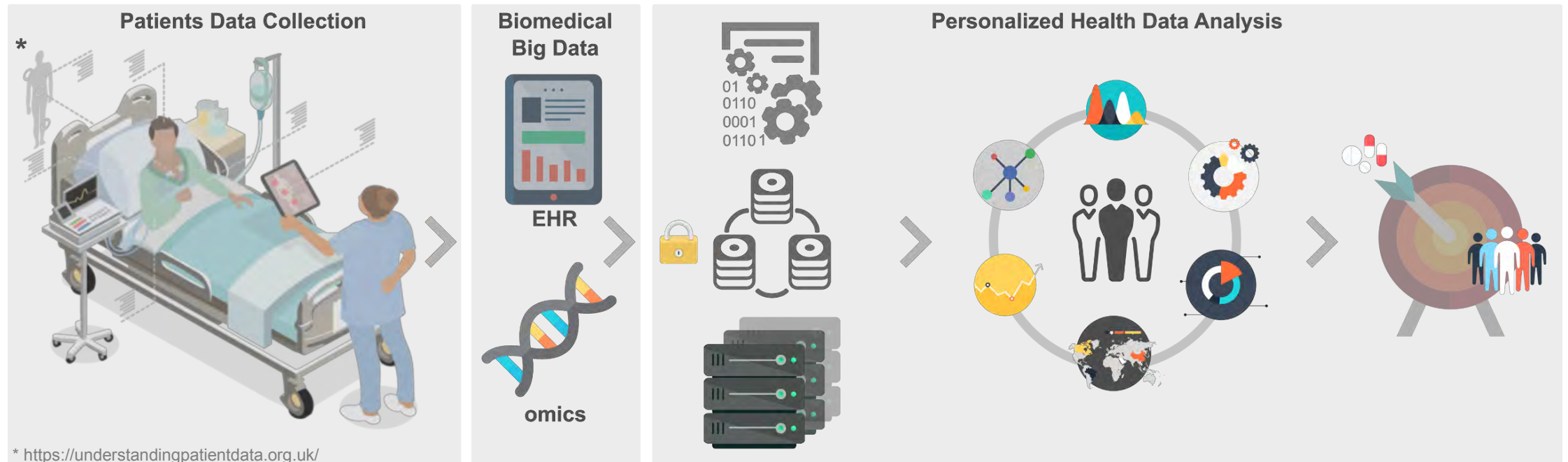
Secure Handling of Confidential Research Data

- Sensitive personal data & Confidential research data & Regulatory frameworks
- Leonhard Med: secure scientific data & IT platform for research on confidential data
- Best practices for secure handling of confidential research data

Sensitive personal data & Confidential research data & Regulatory frameworks

Use case: Personalized Health Research

Provide the **right treatment**, at the **right moment** to the **right patients** (precision medicine) and ensure that as many people as possible **stay healthy** (prevention; personalized health).



○ Research on human data: **sensitive personal data**
Hereafter, per convention: **confidential research data**

Sensitive personal data & Confidential research data & Regulatory frameworks

Use case: Personalized Health Research **Trends**

Confidential research data

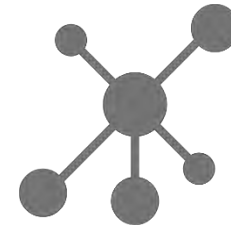
Individual Studies



Classical Biomedical Research
Data Driven Biomedical Research

Now

Networks of Studies



Citizens



<https://www.laboursolutions.com.au>

Data FAIRness

Priv.
Conf.
Risk

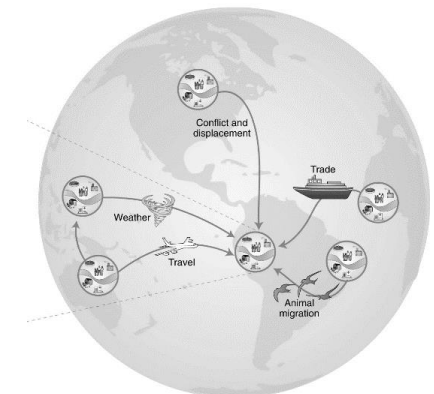
Datasets are combined

One Health



<http://www.oie.int/en/for-the-media/onehealth/>

Global Health

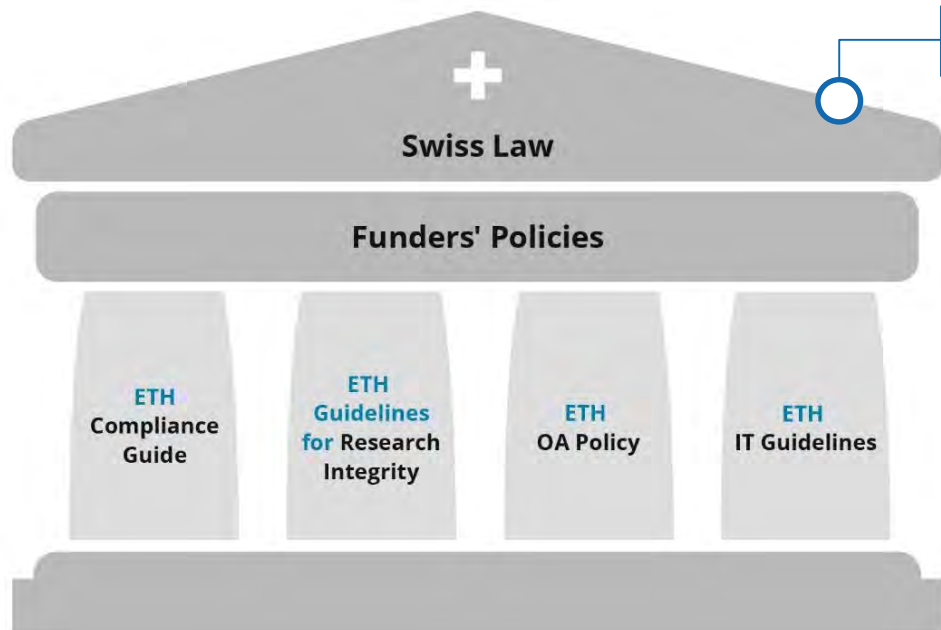


Hernando-Amado et al., 2019, Collignon et al., 2018

Sensitive personal data & Confidential research data & Regulatory frameworks

Confidential Research Data

○ Sensitive Personal Data, *FADP Art. 3c*



Federal Act on Data Protection
Art. 3c Sensitive Personal Data

Human Research Act
Art. 32-35 Further Use of Biol. Material
and Health-Related Pers. Data for Research

Human Research Ordinance
Art. 25-27 Anonymization, Coding
Art. 28-31 Informed Consent

Swiss Criminal Code
Art. 321 Breach of Professional
Confidentiality (rel. HRA)

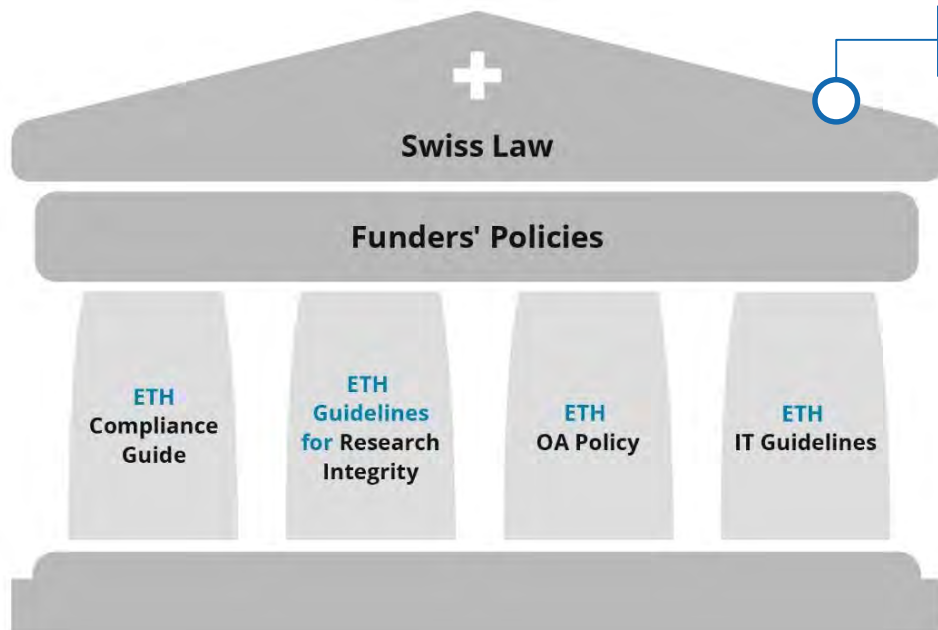
Adapted based on: Matthias Töwe
<http://www.library.ethz.ch/en/Services/Courses-and-guided-tours/Workshops>



Sensitive personal data & Confidential research data & Regulatory frameworks

Confidential Research Data

○ Sensitive Personal Data, *FADP Art. 3c*



Adapted based on: Matthias Töwe
<http://www.library.ethz.ch/en/Services/Courses-and-guided-tours/Workshops>



Federal Act on Data Protection Art. 3c Sensitive Personal Data



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

The following definitions apply:

- a. *personal data (data)*: all information relating to an identified or identifiable person;
- b. *data subjects*: natural or legal persons whose data is processed;
- c. **sensitive personal data**: data on:
 1. religious, ideological, political or trade union-related views or activities,
 2. **health**, the intimate sphere or the racial origin,
- e. **processing**: any operation with personal data, irrespective of the

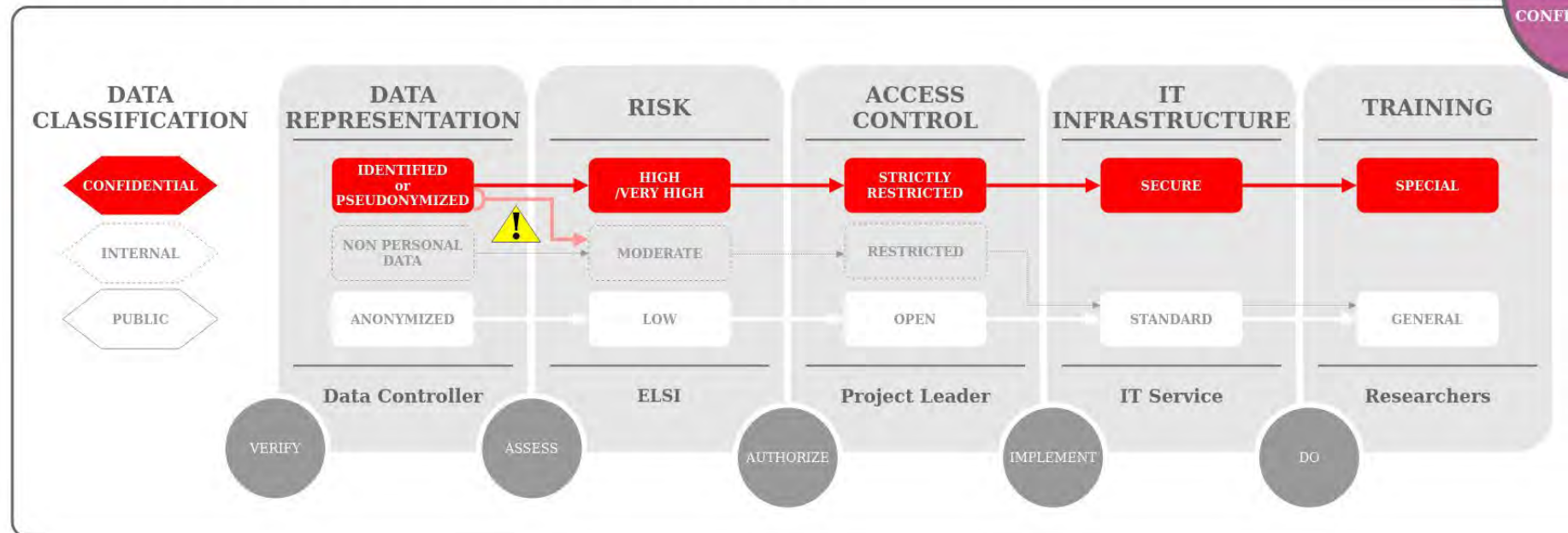
Sensitive personal data & Confidential research data & Regulatory frameworks

YES or NO: **Confidential Research Data** compliant with Swiss legislation and policies

Data classification policy: Leonhard Med, ETH and SPHN *Swiss Personalized Health Network*

All personal data (either identifying data or pseudonymized) are **confidential** *Sensitive Personal Data (FADP Art.3c)*, unless explicitly classified differently.

ENSURE DATA PRIVACY & CONFIDENTIALITY

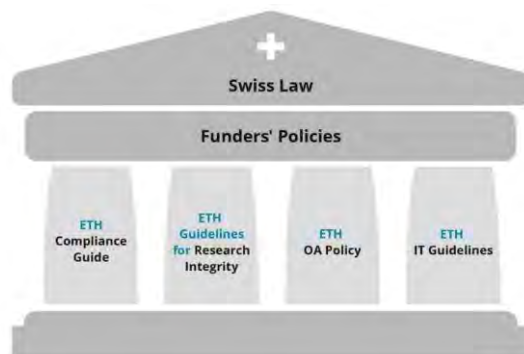


Data classification scenarios <-> Technical & Organizational Measures

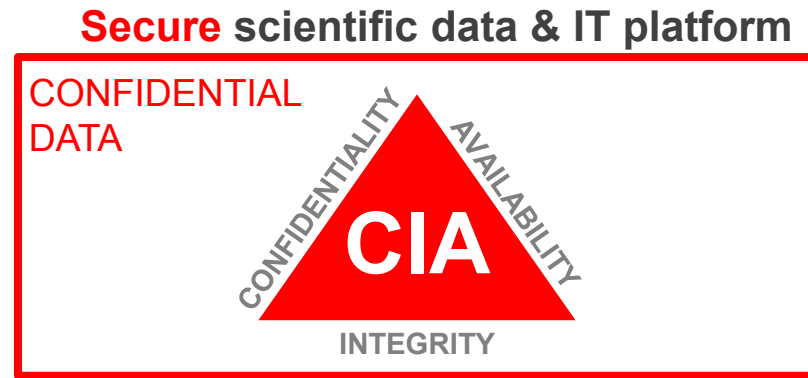
Sensitive personal data & Confidential research data & Regulatory frameworks

Do researchers need to use special secure IT infrastructures?

Yes, if handling **CONFIDENTIAL DATA** *SENSITIVE PERSONAL DATA (FADP)* for research



Adapted based on: Matthias Töwe
<http://www.library.ethz.ch/en/Services/Courses-and-guided-tours/Workshops>



Technical & Organizational Measures



Leonhard Med

OR / AND

infrastructures with similar level of security*

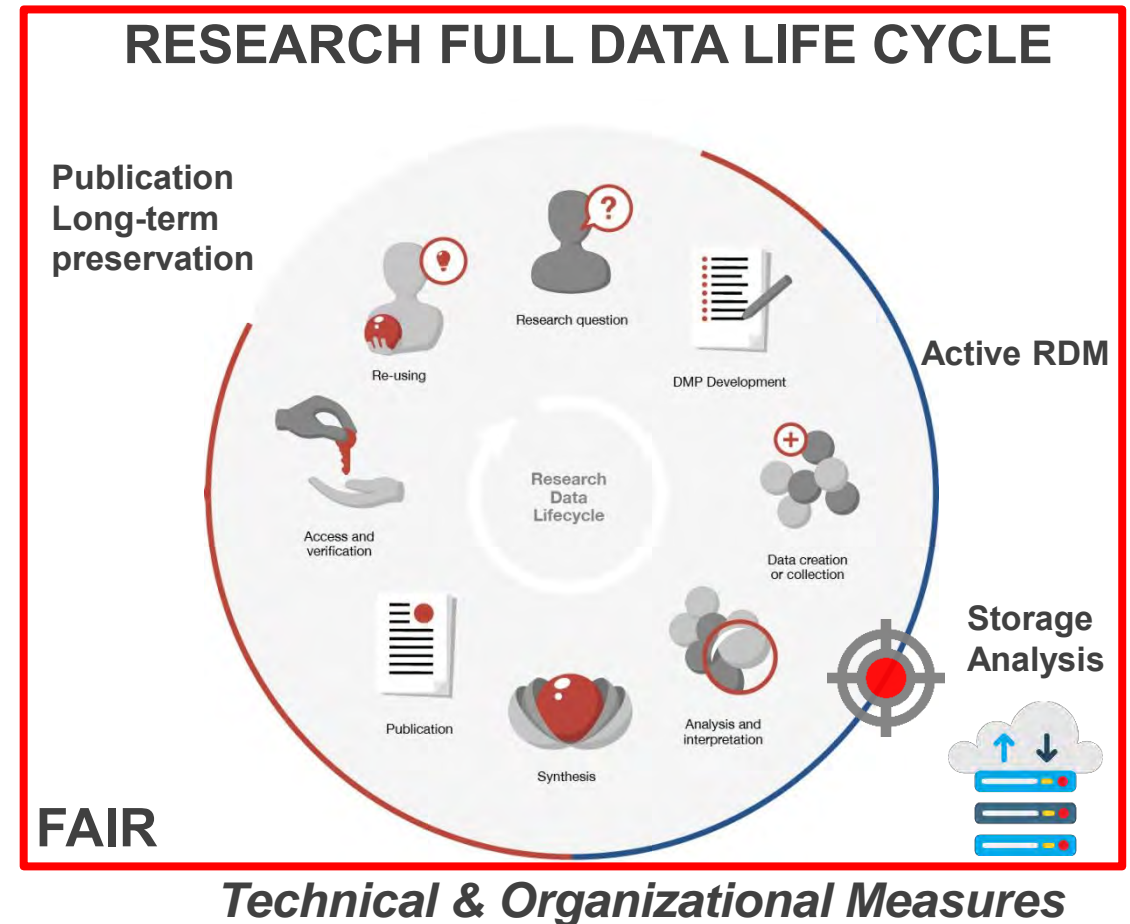
**consult your local IT support*

Secure handling of **confidential research data** - Technical & Organizational Measures (TOM)

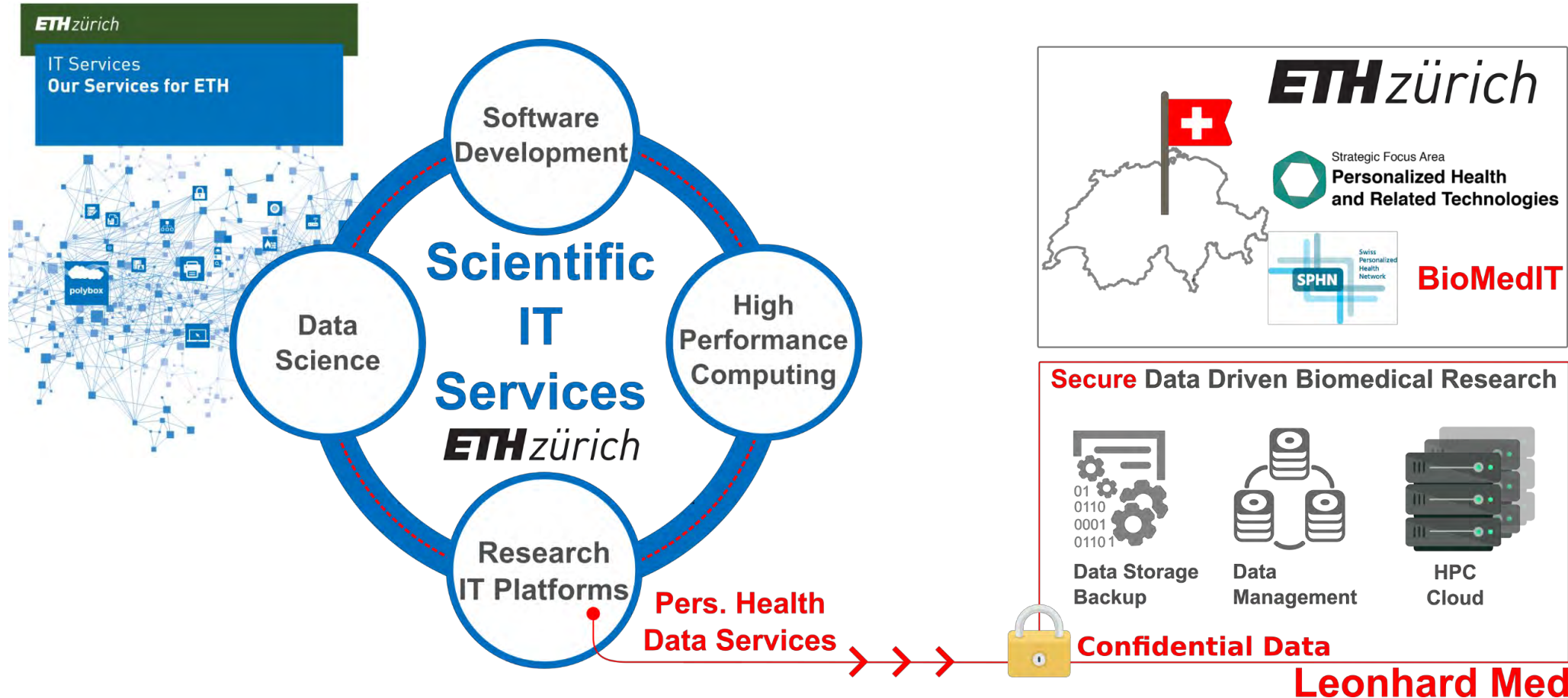
Regulations for confidential research data

- Study participant **informed consent**
- **Ethical approval** to make data available for reuse in research
- **Contract** for data transfer from controllers (e.g. Hospitals) to secure Data & IT infrastructures (processors) where Researchers analyze data
- **Users** (e.g. researchers) are *Responsible* to comply with applicable regulations
- **Project Leaders** (e.g. PIs) are *Accountable* for the research Data Management over the full data life cycle
- **Secure Data and IT infrastructures** (data processor)

Secure Data and IT infrastructures



Leonhard Med: secure scientific data & IT platform, Scientific IT Services at ETH Zurich



Leonhard Med: secure scientific data & IT platform for **bioinformatics** applications in data driven **biomedical research**

<https://sis.id.ethz.ch/services/confidentialresearchdata/>
<https://sis.id.ethz.ch/>
<https://ethz.ch/services/en/it-services.html>

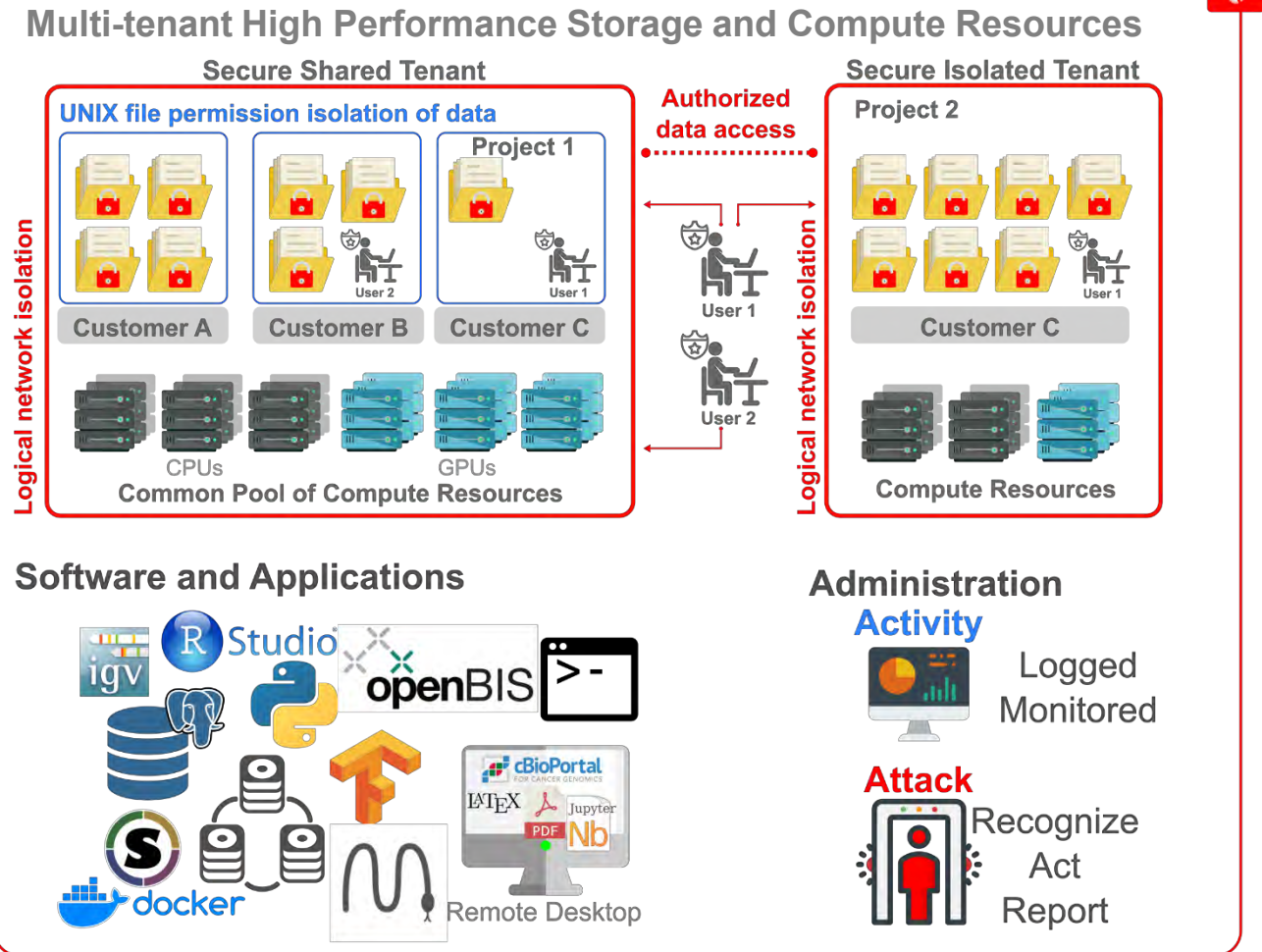
Leonhard Med for ETH and Swiss Research with Confidential Data

Leonhard Med

- is a powerful research IT platform to securely store, manage and process (e.g. bioinformatics, data science) **confidential research data**
- enables collaborative, large-scale and very diverse **biomedical research** (including academies and hospitals) at ETH Zurich
- is part of the national **BioMedIT network** of secure data centers supporting projects in the SPHN and PHRT national programs

SPHN: Swiss Personalized Health Network
PHRT: Personalized Health and Related Technologies

Leonhard Med: secure scientific data & IT platform



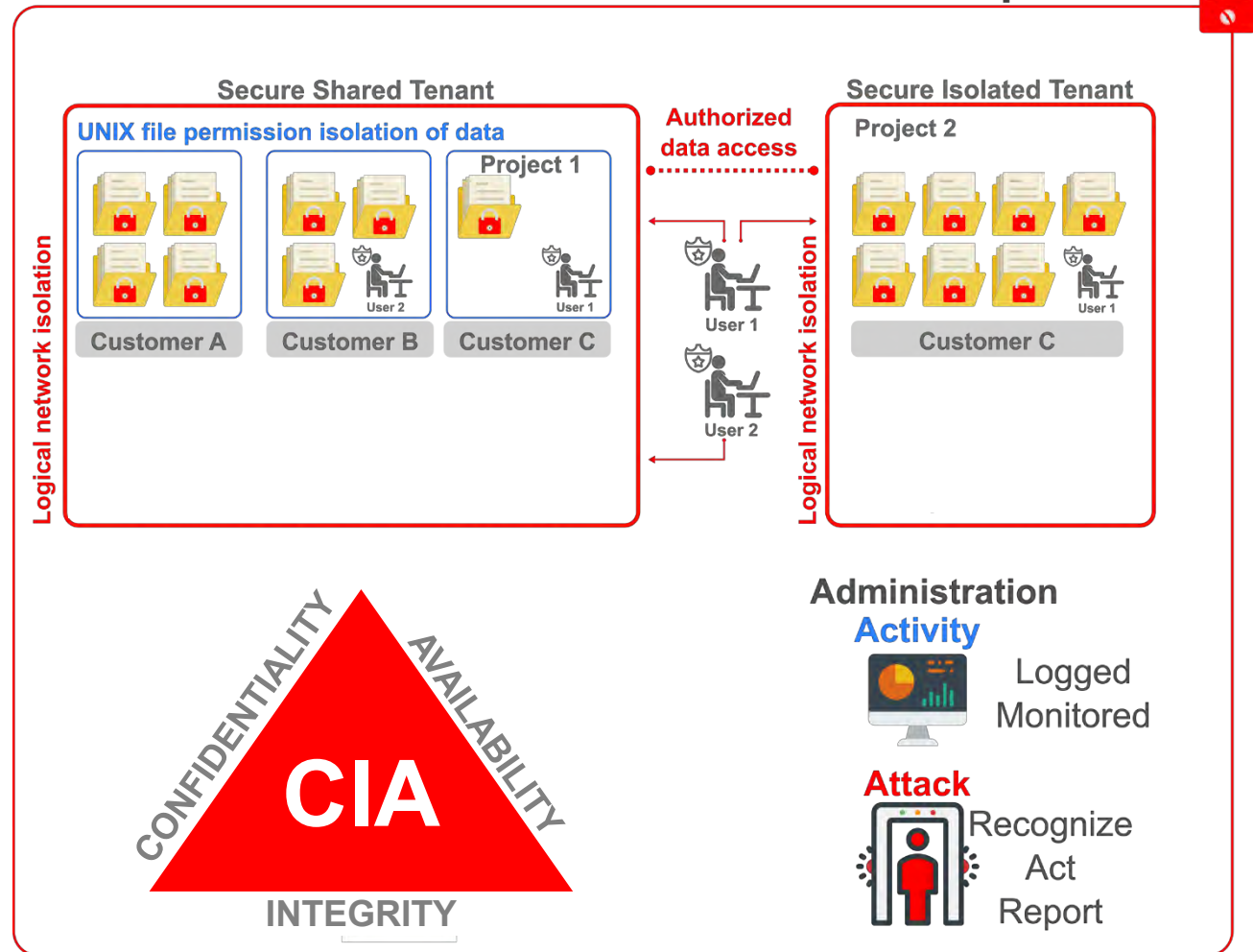
Technical & Organizational Measures

Leonhard Med security controls

CIA measures, confidential data

- physical security of the data center
- data strictly isolated between customers
- user access strictly restricted
- internet access strictly controlled
- data encryption during transfer & in backup
- logging & monitoring
- Leonhard Med Acceptable Use Policy

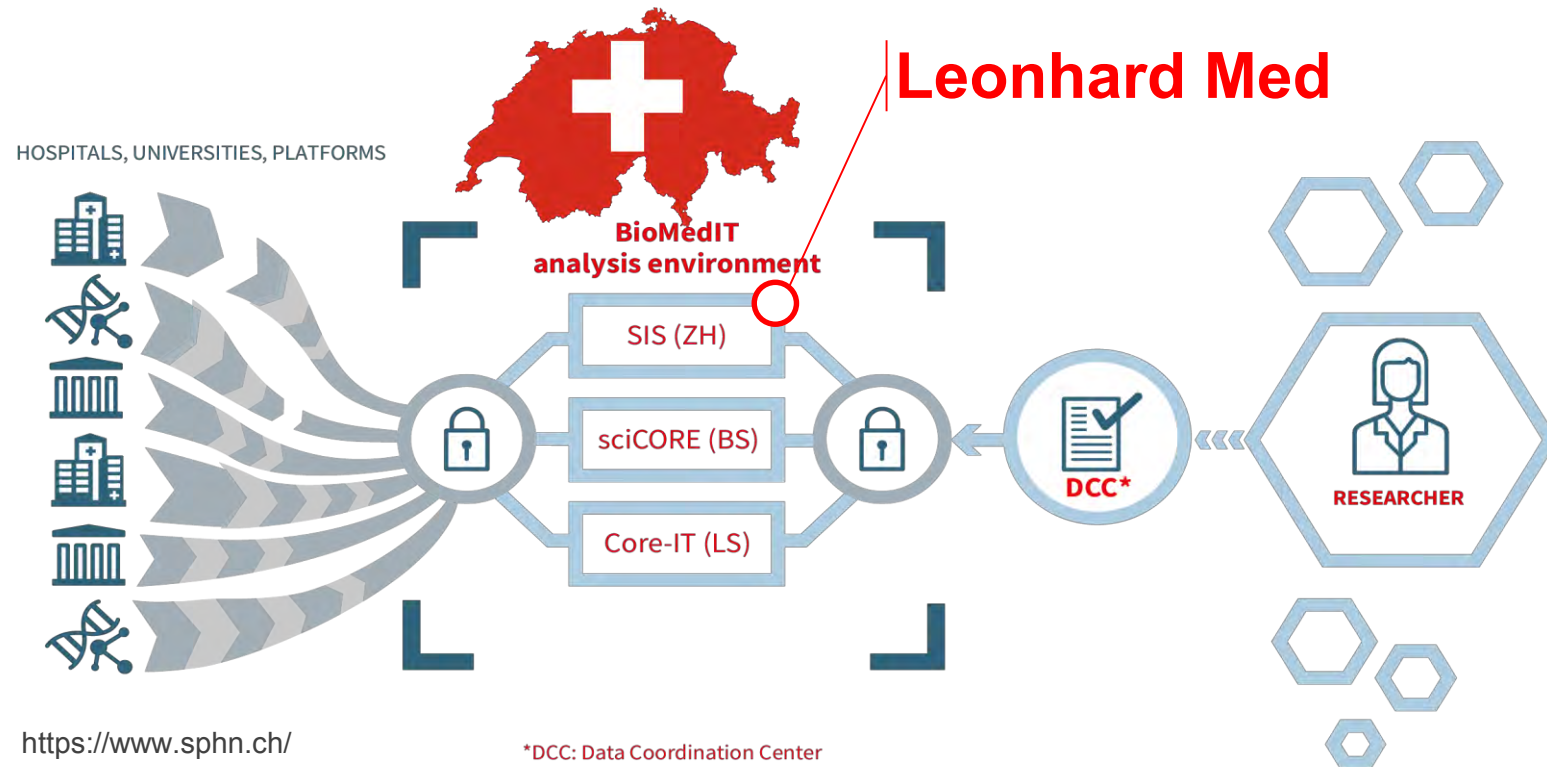
Leonhard Med: secure scientific data & IT platform



Technical & Organizational Measures

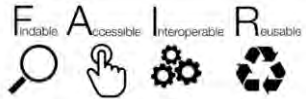
Leonhard Med: the Zurich regional node of **BioMedIT**

Lowering computational boundaries for research with confidential data



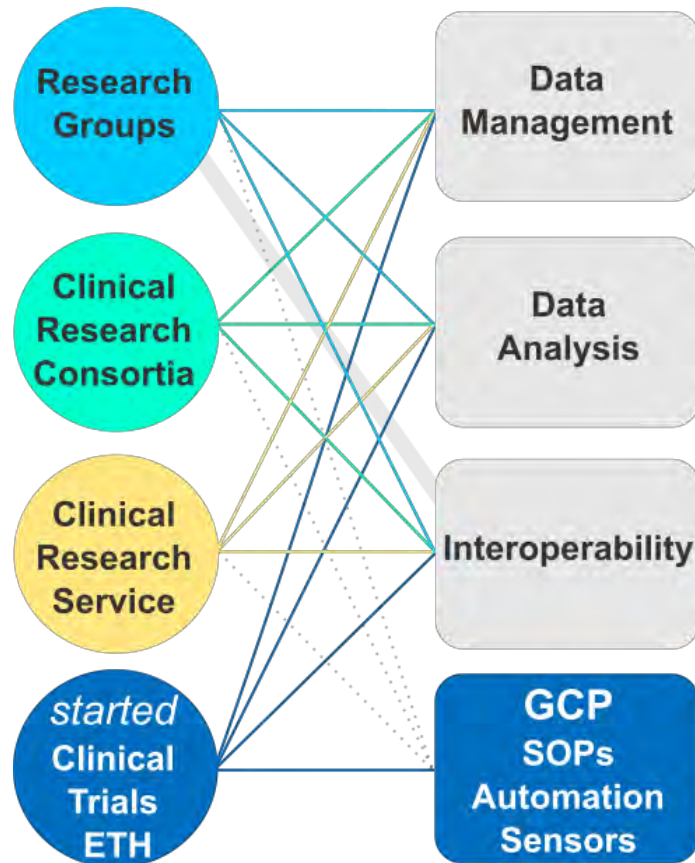
- Data governance
- Secure data transfer process
- Data mgmt., data FAIRness
- Restricted data sharing (secondary use)
- Interoperability of data and bioinformatics workflows
- Data Privacy & IT Security Training

Leonhard Med Use Cases



Research Requirements

Leonhard Med



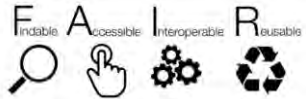
Complete data life cycle, customizable solutions (e.g. secure data transfer, data management, encrypted backup etc.)

Cutting-edge compute: **Bioinformatics, ML, Statistics** (e.g., Bioconductor, R, Python, TensorFlow), Containers, Renku, HPC and Large Volume Data

Data and Workflow interoperability across the BioMedIT network

GCP compliance, QMS, SOPs, bioinformatics and ML workflows for clinical trials at ETH

Leonhard Med as Bioinformatics Platform



Research Requirements

Leonhard Med

Usage & Users

Interinstitutional



Universities
Hospitals
Research Facilities
National Registries
*Citizens

Interdisciplinary



Researchers, Clinicians
IT, Ethics, Law experts
Computer scientists
Molecular biologists
Data scientists

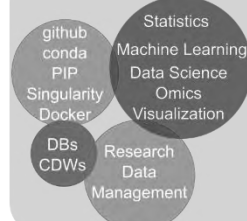
Sustainable



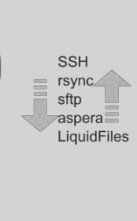
Interoperability of data
Portability of code
Reproducibility
Data mgmt. & sharing
Long term data storage

Data, Tools & Resources

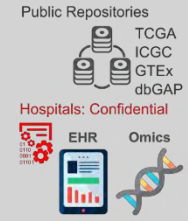
Analytical Software



Data Transfer



Data Sources




Secure access

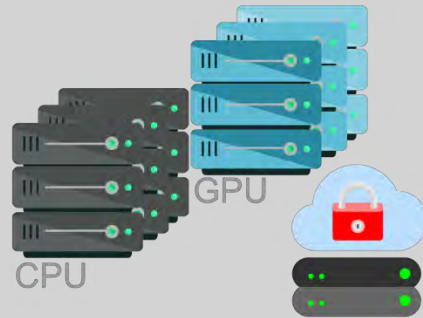


Secure
DataTransfer


Confidential Data




Multi-tenant High Performance Storage




High Performance Computing



Data Analytics and Management Tools



Usability



Administration

Leonhard Med: secure scientific data & IT platform

How can it be used for research?

Leonhard Med data & IT platform

- ✓ secure project space for data storage, management and analysis
- ✓ secure and collaborative compute environment for data analysis (bioinformatics, ML, etc.)
- ✓ secure data transfer from distributed sources
- ▶ customizable e.g., hosting specialized web-based applications, DBs, clinical data warehouses etc.

Leonhard Med – the challenge, the achievement, the future

Secure & Versatile & User-friendly Scientific Data and IT Services

 in production since **2018**

 **25** project spaces

 **> 200** users

 **> 2 PB** data

 + cloud, GCP, web services

Legal & Ethics

- Sensitive personal data
- Data types & legal specification, ethics
- Strictly controlled data access and sharing

Infrastructure & Services

- Distributed and large data lakes
- Data centralization/federation
- Interoperability
- *FAIR* Data
- Compute power (cloud, HPC, federated)

Usage

- Researchers: security & usability
- Clinicians: UX, web services
- Citizens:
<https://www.midata.coop> MIDATA.coop
<https://swissdatacustodian.ch> Swiss Data Custodian

NEW ecosystem

Best practices for secure handling of confidential research data



- Know which data are **confidential** and which access restriction rules apply (i.e., as per DTUA)
- Aim for FAIRness of **confidential research data** (*note: FAIR data may be under restricted access; code & analytic methods under open access*)
- Be “security aware” (i.e., secure your computer) and follow relevant trainings (e.g. SPHN/BioMedIT)
- Follow data privacy rules (laws, policies, guidelines) and only handle **confidential research data** for which you have explicit authorization (i.e., by a Project Leader)
- Use TOMs over the full data life cycle, e.g., secure Data and IT infrastructures, data mgmt. plan, long term preservation etc.)
- Report promptly security incidents to responsible persons at your institution

Best practices for secure handling of confidential research data

○ Don't

- Don't work with data if you don't have explicit authorization, in doubt ask your research supervisor
- Don't share your login account used to access secure systems
- Don't copy **confidential research data** on your computer without authorization (i.e., by PL)
- Don't send **confidential research data** per regular e-mail or tools like regular Dropbox
- Don't share **confidential research data** for which you have use authorization without PL's permission
- Don't consider without formal confirmation (i.e., at Data Provider) that coded/pseudonymized data are anonymized and thus non-confidential
- Don't publish **confidential research data** in public repositories (e.g. GitHub)

Best practices for secure handling of confidential research data

○ Incompliance with relevant policies >> consequences

- Revocation of access right to **confidential research data**
- Revocation of access right to respective secure Data and IT infrastructures (e.g. Leonhard Med)
- Sanctions at home institute, Project Leader is accountable
- Depending on impact (i.e., minor vs. major data breach, unintentional vs. criminal misuse) individual researcher is accountable
- Swiss penal code might be applied
- Research impact (institution reputation, restrict data access etc.)
- *Note: moral obligation towards the data subjects*

Best practices for secure handling of confidential research data

○ | *The 3 top* things to think about *Your* research & confidential data

Data Confidentiality and Individuals Privacy Protection

- ▶ Data **classification**: e.g. public, internal OR confidential
- ▶ Use TOMs for **confidential research data** in your research, e.g. Leonhard Med
- ▶ .. over **the full data life cycle**, i.e., FAIR

Secure handling of confidential research data

Information on Leonhard Med platform and services, regulations and expert consulting


Leonhard Med get in touch

- ✓ **Scientific IT Services, ETH Zurich and services for confidential data**
<https://sis.id.ethz.ch/services/index.html#confidential-research-data>
- ✓ **Leonhard Med Service Level Agreement**
<https://ethz.ch/services/en/it-services/catalogue/server-cluster/hpc.html>
- ✓ **Leonhard Med Acceptable Use Policy**
<https://rechtssammlung.sp.ethz.ch/Dokumente/438.1.pdf>
- ▶ **Leonhard Med custom services and consultancy: contact diana.coman@id.ethz.ch**

Dr. Diana Coman Schmid
Service Manager, Personalized Health Data Services
diana.coman@id.ethz.ch

ETH Zurich
Scientific IT Services
WEC D18
Weinbergstrasse 11
8092 Zurich, Switzerland

<https://sis.id.ethz.ch/>

 @ETH_SIS
@Diana_C_Schmid
@SPHN_ch
@PHRT_CH

Acknowledgements

Swiss Research Data Day 2020

#SRDD2020

22 OCTOBER 2020

Online

- ▶ Leonhard Med Team, SIS ETHZ
- ▶ BioMedIT Team, SIB
- ▶ SPHN, PHRT partner projects