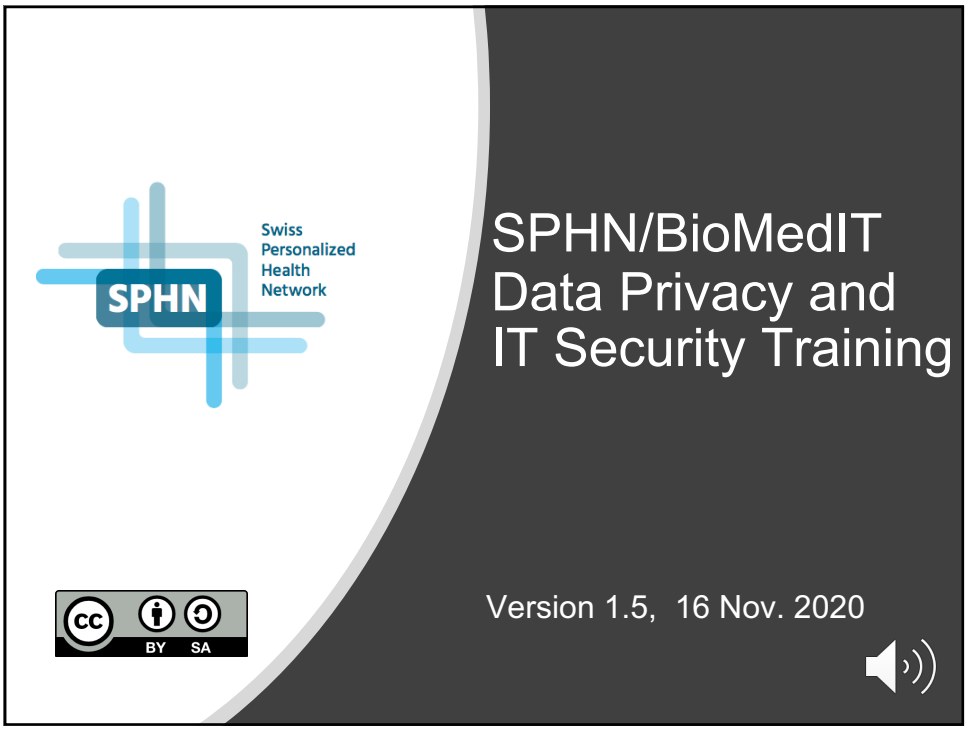
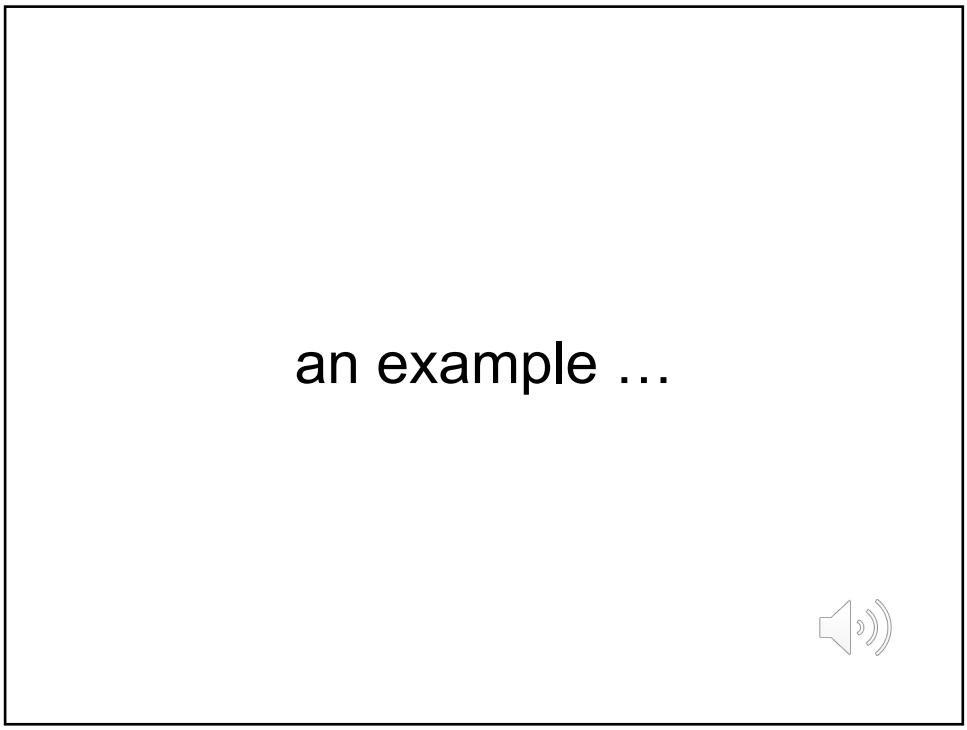
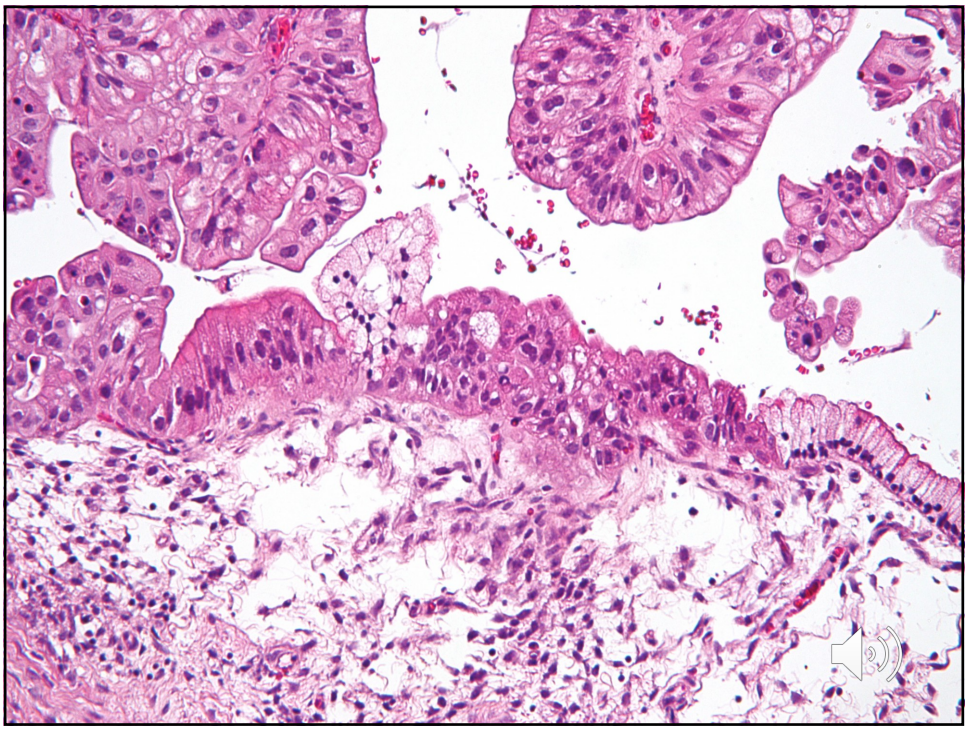

Swiss
Personalized
Health
Network
SPHN
SPHN/BioMedIT
Data Privacy and
IT Security Training
Version 1.5, 16 Nov. 2020

1

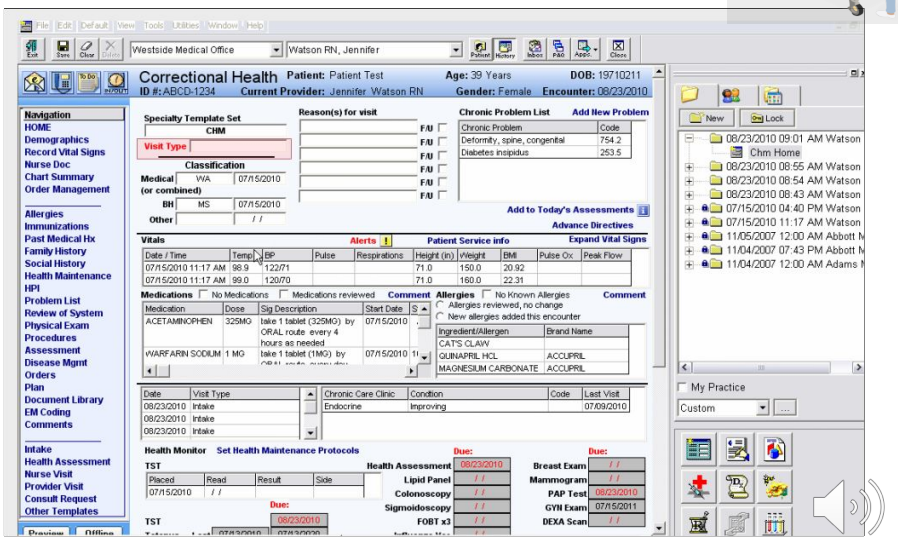

an example …

2



3

Electronic Health Record



https://emr-matrix.org/2010/11/nextgen-healthcare-ehr-frontiers-corrections/

https://understandingpatientdata.org.uk/

4

Issues

Did the patients give consent?

Can’t “just” get patient data

Can’t “just” transfer data in clear text

5

Objective of this training

Train researchers who use **sensitive human** data for research projects

obligatory for SPHN.ch projects and users

6

Objective for participants

Become SPHN-certified to

use human data on IT infrastructures

following regulatory obligations

7

We expect you

To be aware of data privacy and IT security

To adapt your working practice

To apply specific procedures

8

Related work

Complements

Good Clinical Practice (GCP)

9

Target Audience

Researchers (“Users”)

Project Leaders

IT Personnel (BioMedIT Node)

10

Exam for SPHN “Users”

Preparation for “SPHN-certification”

On-line exam to be done within a week

You will receive copies of slides

11

Outline

Data privacy and protection
Laws
Data Classification
IT Security
SPHN - BioMedIT Infrastructure
Rights & Obligations

12

13

Why worry about access restrictions?

14


data
privacy
personal
new
information
identity
social
people
decentralized
internet
digital
one
trust
ecosystem
different
also
web
control
network
used
government
private
systems
rights
service
security
use
communication
public
freedombox
individuals
companies
system
PDJ #3, Spring 2012
cc by Personal Data Journal
http://pde.cc/journal

15

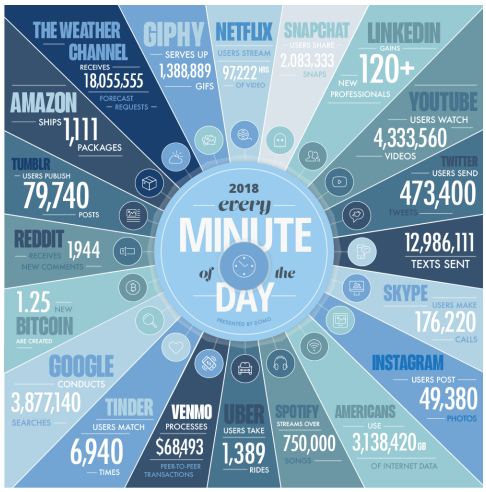
16

Data: personal and sensitive personal data

2.5 quintillion
bytes of data
are created every day

90 % of all data
was produced in the last two years

50 TB of data
will be generated per person every year



https://www.domo.com/learn/data-never-sleeps-6

17

Digital life: do you have control?

Terms of Service

1 HOUR
reading

click & agree

https://www.dimayarovinsky.com/#/i-agree/

Where is my data

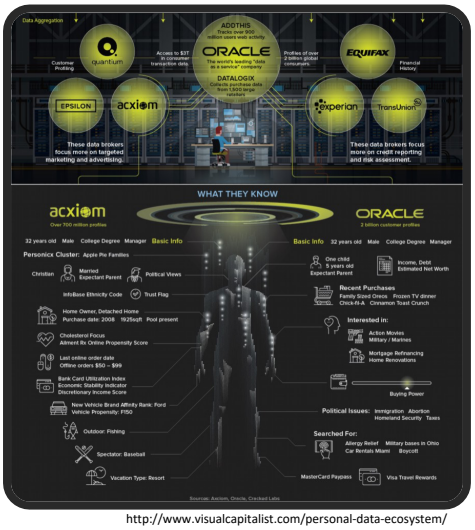
Who uses it

For what

18

Digital life: do you have control?

2.7 billion
personal profiles stored



http://www.visualcapitalist.com/personal-data-ecosystem/

2-4 breaches
of medical data every week (US)

"The Multi-Billion Dollar Industry That Makes Its Living From **Your Data**"

19

Public vs. restricted Data Access

Personal Data Access

Personal Data Footprint

Social media — May be used by anyone — **PUBLIC**

Banking — Security & Privacy for access, storage, processing — **RESTRICTED**

Medical Records — **PUBLICRESTRICTED**
Is my gut bacteria data (microbiome) Patient or Bacterial data?
Persistent data (me, my kids, my relatives)

20

Portuguese Data Protection Authority Imposes 400,000 € Fine on Hospital

The Barreiro Hospital in Portugal was fined 400,000 € by the Portuguese Data Protection Authority CNPD (Comissão Nacional de Proteção de Dados) for in-compliancy with the EU General Data Protection Regulation (GDPR) by not se-parating access rights to patents' clinical data.

Source: https://www.datenschutz-notizen.de/portuguese-data-protection-authority-imposes-400000-e-fine-on-hospital-4821441/

21

The Telegraph HOME | NEWS | SP

News

UK | World | Politics | Science | Education | Health | Brexit | Royals | Investiga

News

Hospital staff disciplined after Ed Sheeran data breach

share Save 4

Ipswich hospital said that staff members "accessed patient information without legitimate or clinical reason" CREDIT: JO HALE/ REDFERNS

Source: https://www.telegraph.co.uk/news/2018/05/19/hospital-staff-disciplined-ed-sheeran-data-breach/

One member of hospital staff has been sacked and another has been given a written warning for accessing Ed Sheeran's personal details without authorisation, it has emerged.

22

May 16 2018

More than 200,000 patients' records were exposed on MedEvolve's public FTP server – researcher

Posted by Dissent at 10:15 am
Breach Incidents, Commentaries and Analyses, Exposure, Health Data, Subcontractor

The researcher who reported the leak to DataBreaches.net observed that a number of clients had files on the FTP server, and in all cases but two, the files were password-protected.

Source: https://www.databreaches.net/more-than-200000-patients-records-were-exposed-on-medevolves-public-ftp-server-researcher/

23

Outline

Data privacy and protection

Laws

Data Classification

IT Security

SPHN - BioMedIT Infrastructure

Rights & Obligations

24

25

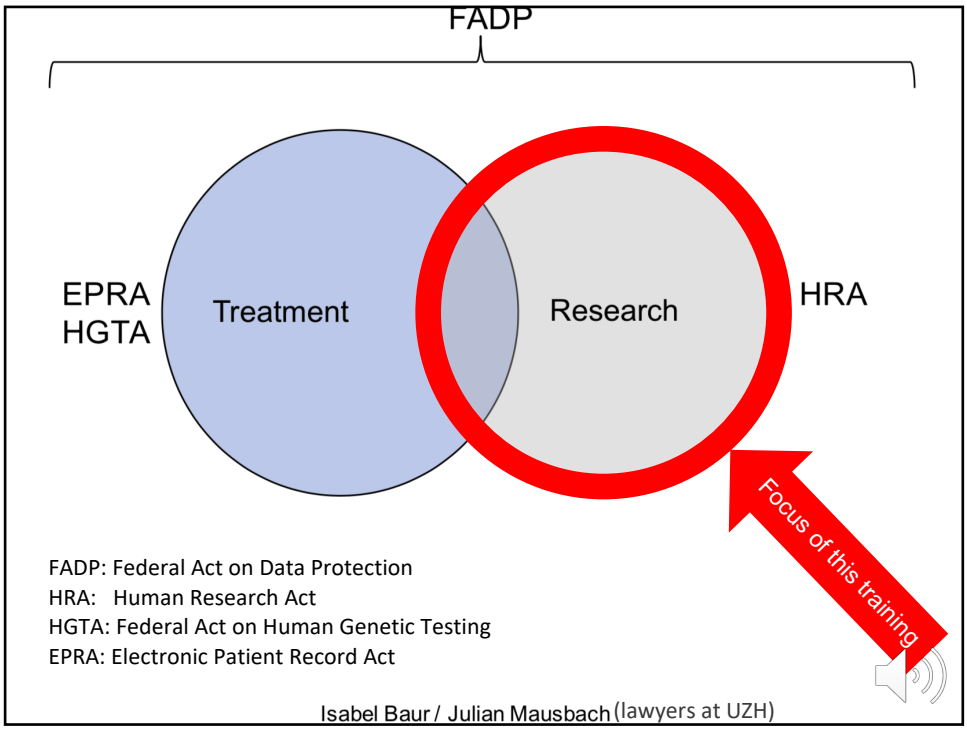
Laws in Switzerland

Federal Act on Data Protection
- Federal Data Protection Ordinance

Human Research Act
- Human Research Ordinance

Swiss Penal Code

26



27

Personal data (FADP, art. 3a)

all **information** relating to an **identified or identifiable person**

28

Sensitive personal data (FADP, art. 3c)

(i) religious, ideological, political or trade union-related views or activities;

(ii) **health**, the intimate sphere or the racial origin;

(iii) social security measures; or

(iv) administrative or criminal proceedings and sanctions

29

Examples of personal data

name

address, phone, email

birth date/place

ID number

biometric information (incl. finger prints)

genetic information

Further reading: **18 HIPAA identifiers** https://en.wikipedia.org/wiki/Protected_health_information

30

Data that can help to identify a person

gender
job position
IP address
hair colour
blood sugar levels
daily movements

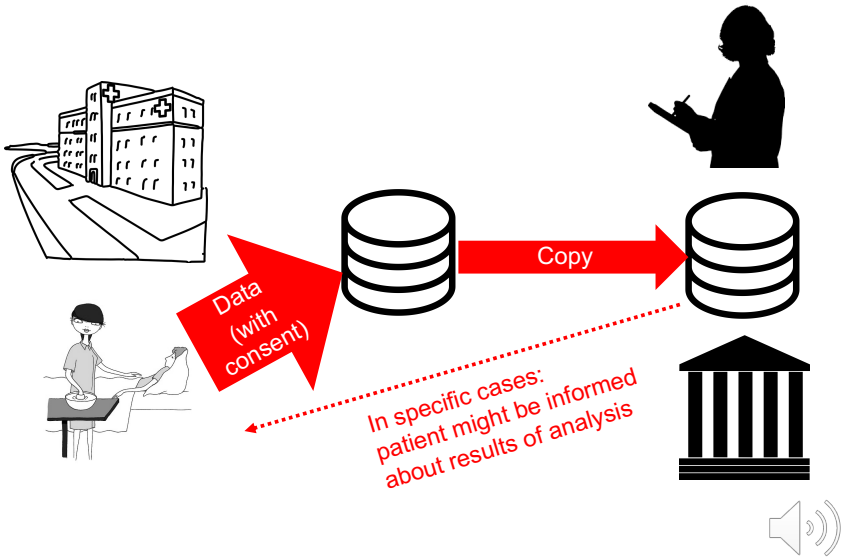
31

Outline

Data privacy and protection
Laws
Data Classification
IT Security
SPHN - BioMedIT Infrastructure
Rights & Obligations

32

Data flow: from hospital to researcher



33

3 important aspects

Patient **consent**

Ethical approval to make data available for reuse

Contract to transfer/use data

34

35

Data classification: 3 categories

Confidential

Public

Internal

Defined by a specific policy – not by the Swiss law

36

Confidential data

Health data (on mobile health devices, apps)

Clinical data (anything that is recorded in a hospital: blood samples, diagnosis reports, family history, microbiome, DNA sequence, etc.)

All personal data
(either identifying data or pseudonymized)
are confidential
unless explicitly classified differently.

37

Confidential data: examples

Genotyping human data (DNA sequencing)

- WGS (Whole Genome Sequencing) data
- WES (Whole Exome Sequencing) data
- Specialized genomic panels (e.g. cancer panels)
- Single cell sequencing, CHiPseq, ATAC-seq
- Information in some types of QC files

Examples: the content of

- raw sequence reads in **FASTA** or similar formats
- **VCF** (Variant Calling Format)
- **SAM** (Sequence Alignment/Map) and **BAM** (binary version of SAM)

38

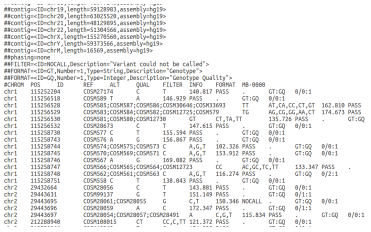
Confidential data in biology & health. guideline only!

- (Transcriptome)
- Patient derived human cell lines
- Proteomics data
- Metabolomics
- Biobanks and cohorts
- Imaging data
- Immunological screening
- Controlled access data in TCGA, dbGAP

39

Confidential data: individual data sets

Genomics



Clinical

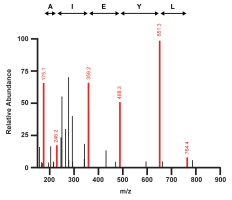
research_ID	gender	year_of_birth	consent_type	consent signed	diagnosis code	procedure	value	unit
USZmU23a1I hxuq6z	male	1973	General Consent	2017-04-21	J77.1	body_weight	74.1	Kg
USZmU28a1I haug2z	female	1971	General Consent	2017-06-02	J17.1	body_weight	68	Kg
USZmU03a2I hxuq6z	male	1961	General Consent	2017-04-21	J37.1	body_weight	89.3	Kg
USZmU23a1I hxuq6z	male	1973	General Consent	2017-04-21	J77.1	body_weight	74.1	Kg
USZmU66a1I hxuq3z	male	1981	General Consent	2017-04-21	J73.1	body_weight	55	Kg
USZmU53a2I hxuq6z	female	1975	General Consent	2017-04-21	J77.1	body_weight	54	Kg

Transcriptomics

mRNA	logFC	FDR	Tumor	Normal
FN1	2.88	1.296151e-19	347787.48	41234.12
COL1A1	1.77	1.680844e-08	358010.32	89293.72
C4orf7	5.20	2.826474e-50	87821.36	2132.76
COL1A2	1.40	9.480478e-06	273385.44	91241.32
GAPDH	1.32	3.290678e-05	179057.44	63663.00
CLEC3A	6.79	7.971002e-74	27257.16	259.60
IGFBP5	1.24	1.060717e-04	128186.88	53323.12
CPB1	4.27	3.044021e-37	37001.76	2637.72
CARTPT	6.72	1.023371e-72	21700.96	215.16
DCD	7.26	1.047988e-80	19941.20	84.80

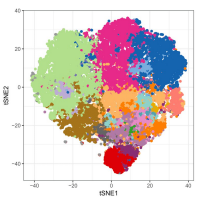
https://bioconductor.org/packages/devel/bioc/vignettes/TCGAbiolinks/inst/doc/analysis.html

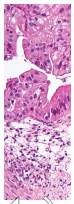
Proteomics



https://www.biozone.utoronto.ca/proteomics-3/

and others





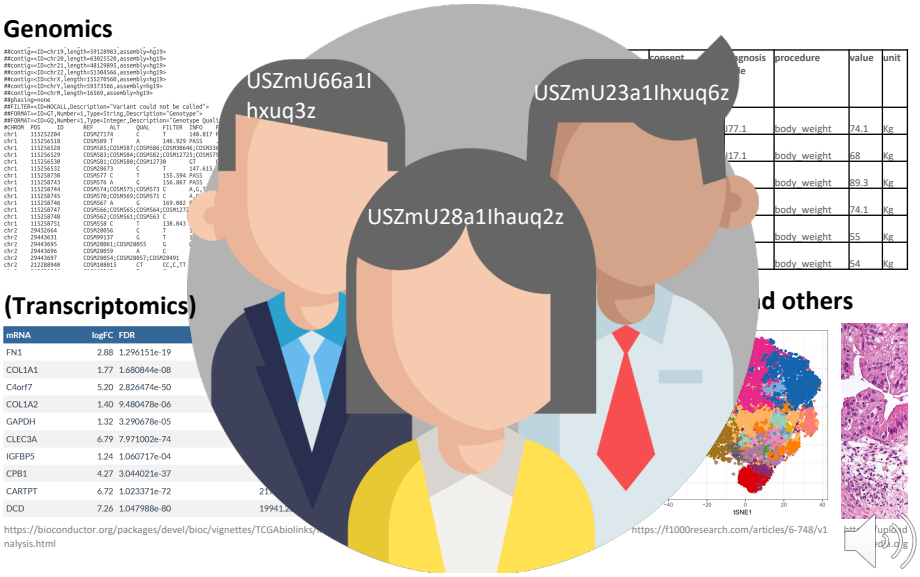
https://f1000research.com/articles/6-748/v1

https://kidneyhealth.org

40

Confidential data: individual data sets

Genomics



(Transcriptomics)

mRNA	logFC	FDR
FN1	2.88	1.296151e-19
COL1A1	1.77	1.680844e-08
C4orf7	5.20	2.826474e-50
COL1A2	1.40	9.480478e-06
GAPDH	1.32	3.290678e-05
CLEC3A	6.79	7.971002e-74
IGFBP5	1.24	1.060717e-04
CPB1	4.27	3.044021e-37
CARTPT	6.72	1.023371e-72
DCD	7.26	1.047988e-80

...d others

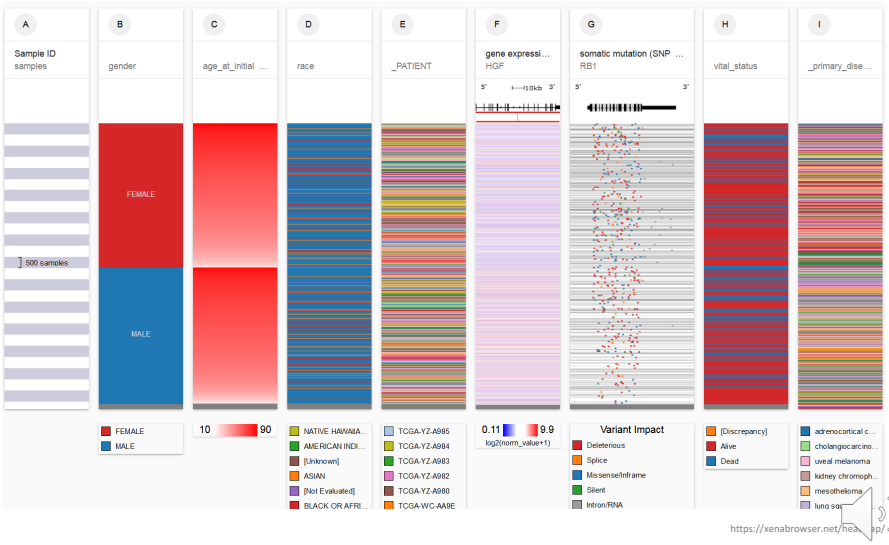
https://bioconductor.org/packages/devel/bioc/vignettes/TCGAbiolinks/...nalysis.html

https://f1000research.com/articles/6-748/v1

41

Confidential data: aggregated data

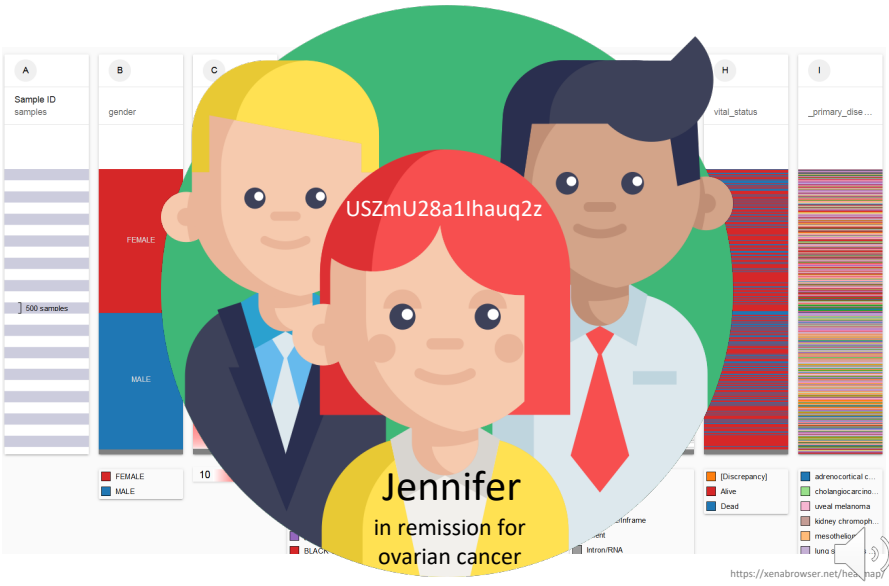
Genomics + Clinical + Transcriptomics + Imaging + ... + Other



https://xenabrowser.net/hea...ap/

42

Aggregated data: higher risk for illegitimate re-identification



43

Public data

Data in various public knowledge bases and archives

Examples

- 1000 Genomes, ExAC, PDB, UniProtKB, etc.
- open access data in: TCGA, ICGC, SRA, genome-phenome archives, etc.

44

Internal data: example

Name of all users involved in SPHN projects

Usually, a policy defines what is "internal"

45

Making data available: data provider

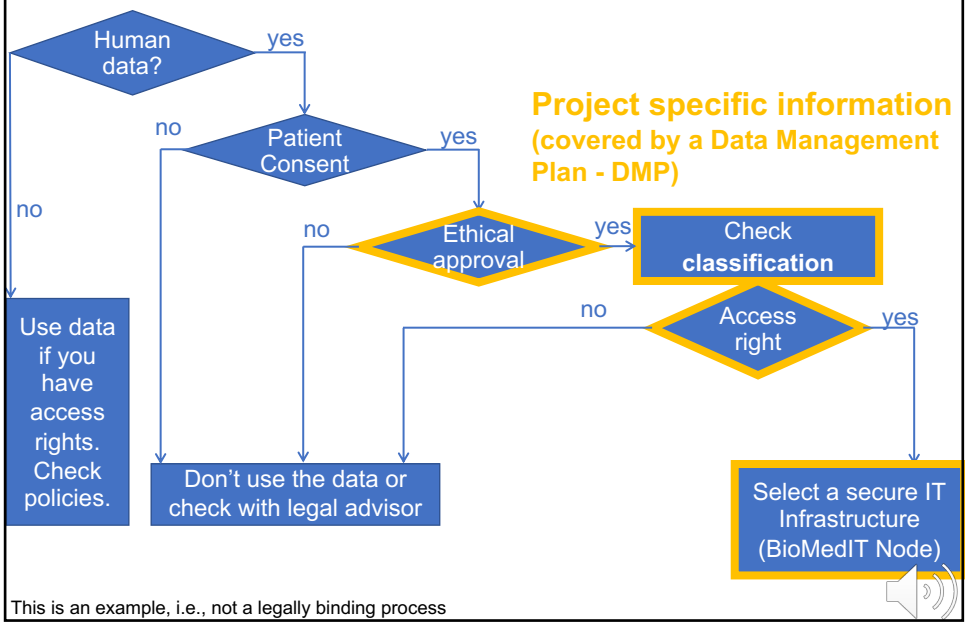
Collect consent

Filter, de-identify, package data
(FAIR principles)

Classify data as confidential

46

Decision tree for data re-use in research



Project specific information
(covered by a Data Management Plan - DMP)

This is an example, i.e., not a legally binding process

47

Representation of data: examples

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Adapted from: **A Guide for technical and organizational measures**
The Federal Data Protection and Information Commissioner (FDPIC) 2015

Identifying

Oscar Smith
13.03.1973

Broke his left leg while skiing

Coded/Pseudonymized

USZmU23a1Ihxuq6z

Broke his left leg while skiing

Anonymized

Someone

Broke his left leg while skiing

Concordance table

Code	First Name, Last Name
USZmU23a1Ihxuq6z	Oscar, Smith 13.03.1973

Encrypted

EnCt2536869affba4 9344ef01b24c2d131 1782ae6a88153686 9affba49344ef01b2 4ch/F6hVeLgQC dzJeg7lt0PHsLwbgg XtcualxRTNOTduyn8 iCJVem/1loEuPtkCQ ==IwEmS

48

Anonymisation, HRO (art. 25)

For the anonymisation of biological material and health-related personal data, **all items which, when combined, would enable the data subject to be identified without disproportionate effort, must be irreversibly masked or deleted**.

In particular, the name, address, date of birth and unique identification numbers must be masked or deleted.

49

Potential risks with data

“Free text fields” can also contain personal data

e.g. patient IDs, sample IDs

50

Research with

anonymized data is

not subject to HRA

51

Coding or Pseudonymization

Identifying information is

replaced by a code (e.g. concordance table)

only accessible with a **"key"** under strict security regulations

52

Pseudonymized data example

⚠ **pseudonymized data is confidential**

data_provider	patient_research_ID	gender	year_of_birth	consent_type	consent signed	diagnosis code	procedure	value	unit	date	system
USZ	USZmU23a1lhxuq6z	male	1973	General Consent	2017-04-21	J77.1	body_weight	74.1	Kg	2018-01-17	KIS
USZ	USZmU28a1lhauq2z	female	1972	General Consent	2017-06-02	J17.1	body_weight	66	Kg	2018-01-17	KIS
USZ	USZmU03a2lhxuq6z	male	1961	General Consent	2017-04-21	J37.1	body_weight	89.3	Kg	2018-01-17	KIS
USZ	USZmU23a1lhxuq6z	male	1973	General Consent	2017-04-21	J77.1	body_weight	74.1	Kg	2018-01-17	KIS
USZ	USZmU66a1lhxuq3z	male	1981	General Consent	2017-04-21	J73.1	body_weight	55	Kg	2018-01-17	KIS
USZ	USZmU53a2lhxuq6z	female	1975	General Consent	2017-04-21	J77.1	body_weight	54	Kg	2018-01-17	KIS

Concordance table	
Code	**First Name, Last Name**
USZmU23a1lhxuq6z	Oscar, Smith

kept at USZ

Identifying

Oscar Smith
13.03.1973

Broke his left leg while skiing (J77.1)

53

Conditions for breaking the code

coded or pseudonymized data -> identified data

breaking the code is necessary to avert an immediate **risk to the health** of the person concerned;

legal basis exists for breaking the code;

breaking the code is necessary to **guarantee the rights of the person** concerned, and in particular the right to revoke consent.

Human Research Ordinance, HRO, Art. 27

54

Terminology summary

de-identification: process used to prevent a person's identity from being connected with information, i.e., the identity of a person can't be obtained anymore

- **pseudonymization** (used in context of GDPR): substitutes the identity of a data subject in such a way that additional information is required to *re-identify* the data subject
- **coded:** personal data and human biological material linked to a specific person via a code (cf. SPHN Glossary)
- **anonymization:** irreversibly destroys any way of identifying a data subject. *Note that anonymization must not be confused with pseudonymization!*

re-identification: process of matching de-identified data with publicly available information, or auxiliary data, in order to discover the individual to which the data belongs to.

encryption: processes of encoding a message (cf. SPHN Glossary)

SPHN Glossary: www.sphn.ch

55

Outline

Data privacy and protection

Laws

Data Classification

IT Security

SPHN - BioMedIT Infrastructure

Rights & Obligations

56

Access to confidential data granted

What **technical measures** to follow as an SPHN User?

57

How do you protect your bank account?

58

59

Researchers need

Secure IT infrastructure

Designed for doing research on human data

60

Secure IT Infrastructure

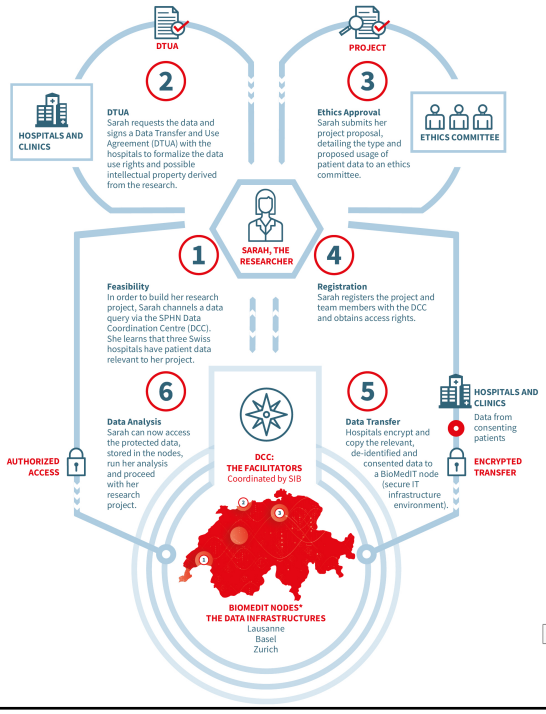
Requires specific practice

Shared with several users

Professional IT Support

61

A use case



62

Outline

Data privacy and protection
Laws
Data Classification
IT Security
SPHN - BioMedIT Infrastructure
Rights & Obligations

63

SPHN - BioMedIT network in Switzerland

Network of secure IT infrastructures

For confidential research data

Specifically supports personalized health research

Fulfils legal and security requirements

64

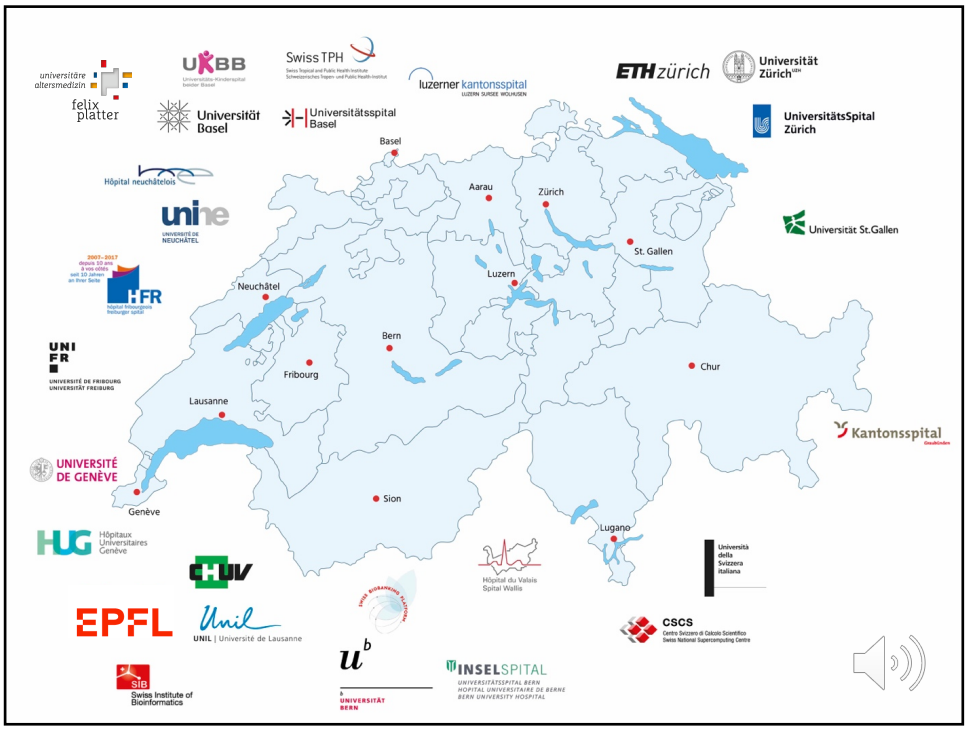
SPHN - Data Coordination Centre

Interoperability of health-related information

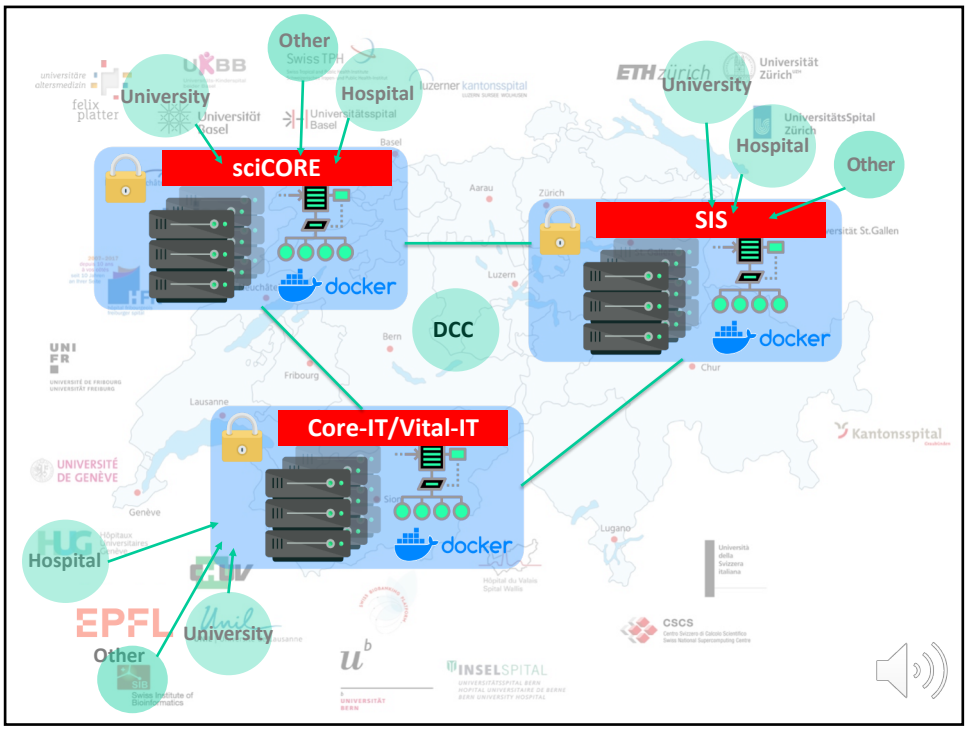
Standards for data formats, semantics, governance, and exchange mechanisms

DCC will **coordinate** user access rights, project registry, monitoring of data usage, etc.

65



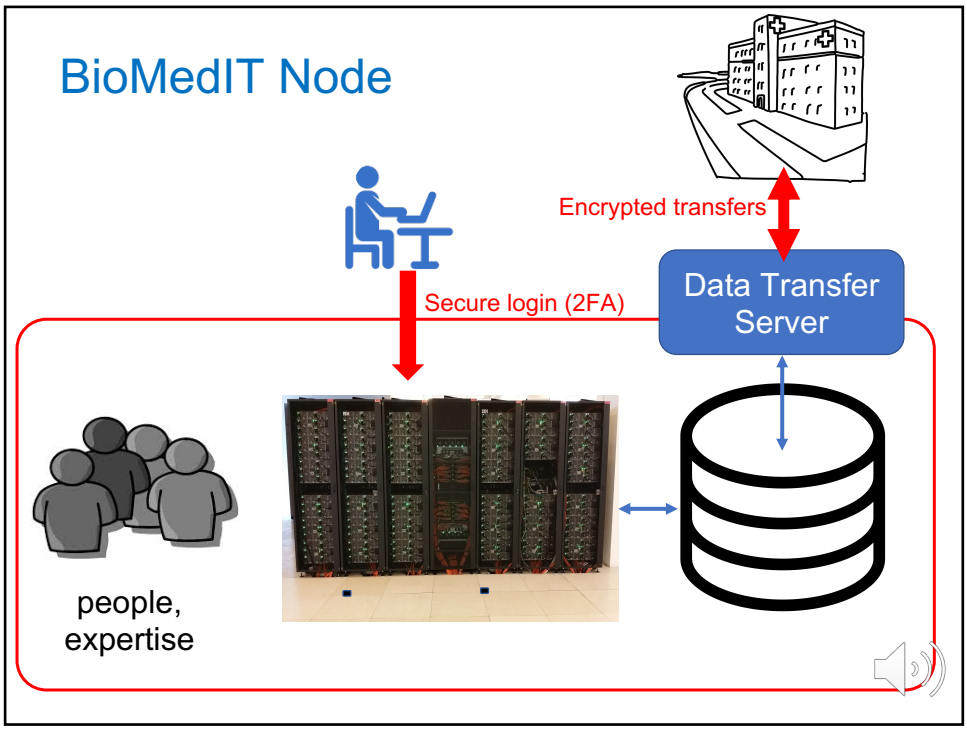
66


Other
University
Hospital
sciCORE
University
Hospital
Other
SIS
docker
docker
DCC
Core-IT/Vital-IT
docker
Hospital
EPFL
University
Other
Kantonsspital

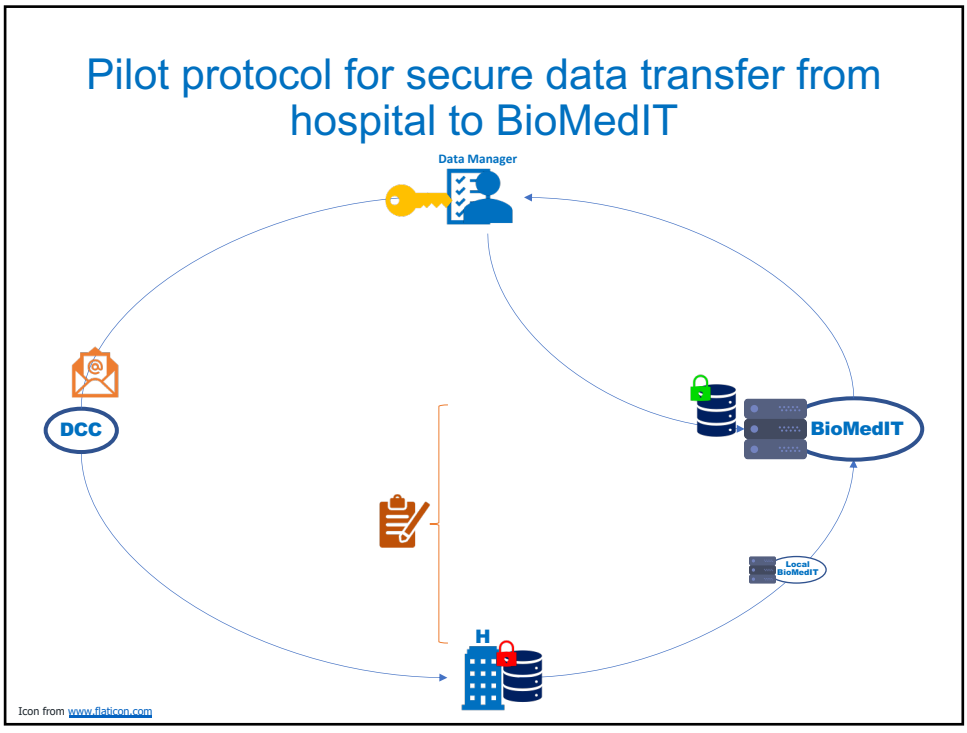
67


people,
expertise

68


BioMedIT Node
Encrypted transfers
Secure login (2FA)
Data Transfer Server
people, expertise

69


Pilot protocol for secure data transfer from hospital to BioMedIT
Data Manager
DCC
BioMedIT
Local BioMedIT
H
Icon from www.flaticon.com

70

Pilot protocol for secure data transfer from hospital to BioMedIT

- Hospital fetches **project encryption key** and verifies validity with DCC
- Hospital **prepares data and metadata** (according to DCC registry, project ID implicitly defines routing)
- Hospital **encrypts data** and sends it to local BioMedIT node
- **Data routing** by BioMedIT nodes
- Data Manager gets notified of delivery to partner (final) BioMedIT node (by DCC, currently by email)
- Data Manager **decrypts data at BioMedIT** node
- **Transfer recorded** in DCC logs

71

Pilot protocol for transferred data package structure

- Should be consistent across hospitals
- Name convention (following ISO 8601 datetime format):
  YYYYMMDDhhmmss
- Data is tarballed and zipped as a tar.gz
- The data is then encrypted
- The metadata file is created
- The metadata file and the encrypted data are then "tarballed" for transfer.

72

Outline

Data privacy and protection
Laws
Data Classification
IT Security
SPHN - BioMedIT Infrastructure
Rights & Obligations

73

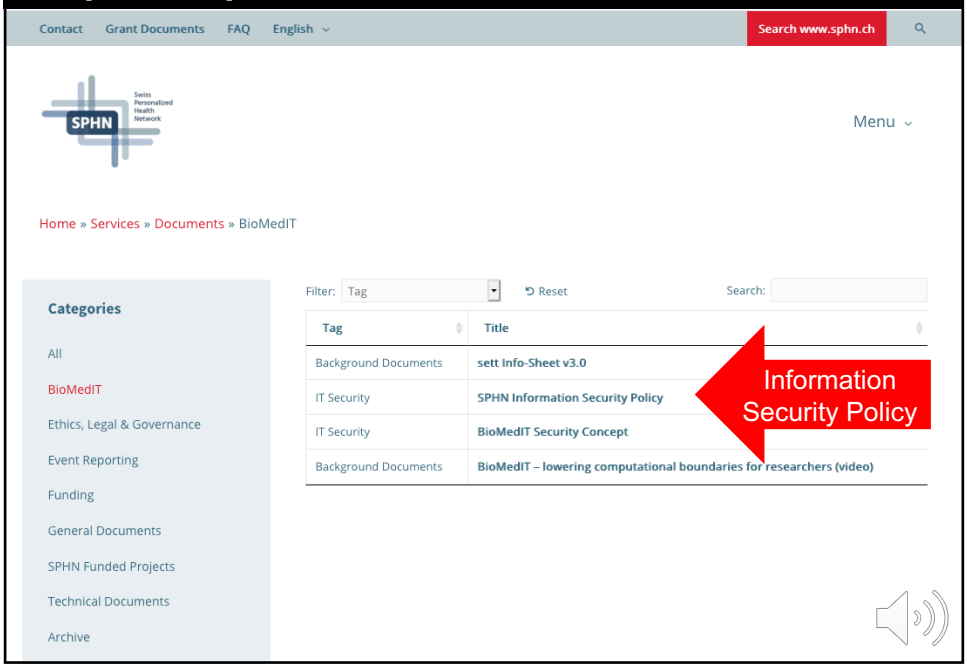
includes link to SPHN ELSI Policy

SPHN/BioMedIT Information Security Policy

Document Number: PL-001
Version: 2.0

74

https://sphn.ch/services/documents/



75

Audience of Info. Security Policy

Project Leaders

Users

BioMedIT Node IT personnel

76

Project Leader & Users

4.2 Classification of Information Assets

5.2 Users and Responsibilities

8 Cryptography

12 Awareness Training

77

BioMedIT Node Personnel

3 IT Security Governance

4 Asset Management

5 Access Control

6 Operations Management

7 Physical and Env…

8 Cryptography

9 Comm…

10

11

12

13

14

15

78

Obligation of Project Leader

Request BioMedIT-access for project and team members

Data life cycle

Project reporting (also in case of issues)

Sign Data Trans. & Use Agreement

79

Data Transfer and Use Agreement

Lists data for SPHN project

Contains project description

Signed by **Project Leader**(s) and **hospital** (data provider)

80

Users: “Acceptable Use Policy”

To be signed by all BioMedIT Users

Covers legal aspects between Users and node

Project Leader has responsibility for project team (Users)

81

User account

Personal: must not be shared

Set good password
- 2 Factor Authentication

Confirm account once/twice a year

Information Security Policy, Section 5.2

82

Data access and sharing

Only use data for which you have **explicit authorisation**

Don't share it with anybody outside your project team

Use **dedicated transfer tools** to import/export data

83

Usage of personal computer

Use latest security patches & software versions

Use malware protection

Disk must be encrypted *if confidential data is used*

84

Encryption

Confidential data transferred into/out of BioMedIT Node must be encrypted

Portable media: encrypted

Various tools are available

85

Disk encryption in brief

Mac: FileVault

Windows: BitLocker

Linux: LUKS

86

Explicitly avoid …

Don’t publish confidential data in public repositories (GitHub, Dropbox, etc.)

87

Non-compliance with policy: possible consequences

Removal of access right to BioMedIT Node

Sanctions in home institute

Danger for BioMedIT

Note: moral obligation towards the data subjects

88

Security incidents: data breach

What to do

Report promptly –
give as much info as possible

- incl. date/time/what data
- risk and impact/consequences

89

Security incidents: data breach

Whom to contact

Project Leader must contact
BioMedIT Node

- sphn.ch/biomedit/

90

Dear BioMedIT Node,

I am the project leader of <PROJECT NAME> where we detected a data breach 1 hour ago.

Three patient records leaked out and were found on the public web site people.myuniversity.edu on 15 Oct 2018. The data are not online anymore.

There is a high risk that the breach will lead to a disclosure of medical information about patients.

Please let me know if you need additional information.

Best regards, Frau Mustermann

91

Summary

- Obtain access rights to data
  - Know about ethical approval, data classification
  - Use a valid Data Use Agreement
- Protect your computer
- Use a secure IT Infrastructure
- Protect your account on BioMedIT
- Don’t share/copy confidential data without permission
- Encrypt confidential data when transferred

92

BioMedIT Node

Access Example

93

How to get access?

Be part of an SPHN project

Sign “acceptable use policy”

More info: sphn.ch/biomedit

94

Access requests

Access requests will soon be possible through the SPHN portal

Until then, direct contact to:

LeonhardMed:	cluster-support@id.ethz.ch
sciCORE:	scicore-admin@unibas.ch
Core-IT/Vital-IT:	it-support@sib.swiss

2 factor authentication

95

BioMedIT Network

Access via web-based remote desktop

ssh access for “power users”

96

Login via SPHN portal

Swiss Personalized Health Network (SPHN)

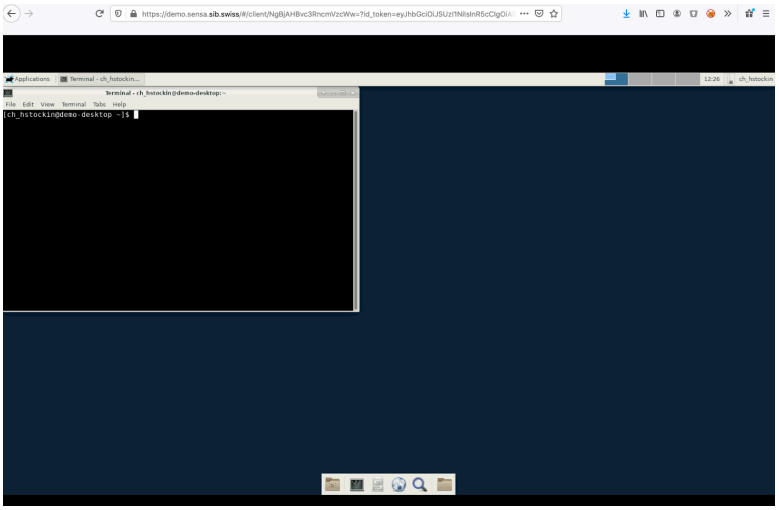
Authentication System

Log In

SWITCHedu-ID

97

BioMedIT Remote desktop



98

Data transfer to BioMedIT Node

Formal appointment of a **Data Manager**

Formal request to open a working space on the responsible BioMedIT node

List of collaborators who will work with project data

99

Conclusion

Data privacy & confidentiality

Access restriction

Technology to guide you

100

Exam

Confirm if you want to do exam

We will send link to online exam

You have 1 week to do exam

101

Thank you.

102

About the speakers

Sofia Georgakopoulou
University of Basel & SIB

Diana Coman Schmid
ETH Zurich & SIB

Heinz Stockinger
SIB Swiss Institute of Bioinformatics

103

Acknowledgements

UniversityHospital Zurich
- Karin Edler
- Francisca Jörger
- Cornelia Kruschel
- Michael Weisskopf

ETH Zurich – SIB
- Christian Bolliger
- Diana Coman Schmid
- André Kahles
- Simona Morello
- Gunnar Rätsch
- Bernd Rinn
- Thomas Wüst

University of Zurich
- Isabel Baur
- Julian Mausbach

University of Basel – SIB
- Sofia Georgakopoulou
- Thierry Sengstag

SPHN/DCC – SIB
- Leila Alexander
- Katrin Crameri
- Martin Fox
- Kevin Sayers
- Silvia Schaub
- Torsten Schwede

SIB Swiss Institute of Bioinformatics
- Séverine Duvaud
- Roberto Fabbretti
- Marc Filliettaz
- Vassilios Ioannidis
- Warren Paulus
- Grégoire Rossier
- Heinz Stockinger

SIB Data Protection & IT Security Board

External Reviewer

And all people who participated in the SPHN Information Security Policy

104