

Data and Ethics Handbook

Selective Perspectives in Autumn 2022



Figure 1 Cybersecurity

Duration: Start 2022-09 till 2022-11

Author: Ekaterina Golubeva

Email: golubeka@students.zhaw.ch

Program: Module D4-Data and Ethics-HLS AS2022

Professors: Dr.Pascal Moriggl and Prof.Dr.Petra Maria Asprion

Management Summary

In this notebook, lessons from « Data and Ethics » lecture given by Professors Pascal Moriggi and Petra Maria Asprion are put into practice, summarized, and discussed.

We first started by exploring Personal Cybersecurity and good practices to protect oneself from cyberattacks and negligent and risky behavior with regards to data. We studied different threats, terminology, measures, and some applications (encryption of messages, installing protective softwares etc.)

We then explored the relevance, history and GRC conditions and policies from the organizational perspective. Information Security was at the core of our interest to know the value of data and to assess risk of cyber-attacks in a company. It was put into practice through an Information Security Policy that we designed for a company of our choice.

Another important aspect of data was data management and FAIR principles, which help us in keeping our data clean, accessible, and reusable. We put our acquired data stewardship knowledge to use by creating a Data Management Plan, a codebook, and a data map.

Finally, we created a Data Ethics Canvas to investigate and access moral, legal, and ethical considerations related to data.

Table of Content

Management Summary	2
Table of Content	3
1 Relevance Description	4
1.1 Awareness	4
1.2 Personal Information Security	5
1.3 Organizational Information Security	6
1.4 Terminology	7
2 Information Security Policy	8
2.1 Definitions of information security, terms	8
2.2 Own perspective	9
2.3 Organizational perspective	9
2.4 Additional thoughts	10
3 Data Management Plan	11
3.1 Project scenario	11
3.2 Data Perspective and DMP description	12
4 Fair Guiding Principles	14
4.1 Organizing data	14
4.2 Metadata structure	15
4.3 Codebook for example data set	15
4.4 FAIR elements for example dataset	16
5 Data Ethics Canvas	17
5.1 Ethics terminology	17
5.2 Ethics Review	18
5.3 Reflection on Data Ethics Canvas	18
6 Conclusion and Reflection	19
6.1 Content	19
6.2 Coaching Sessions	19
6.3 Personal view	19
7 Bibliography	20
8 List of Figures	21
9 Appendix	22
9.1 Information Security Policy	22
9.2 Coaching session 2: Cybersecurity	23
9.3 Data Management Plan	24
9.4 Codebook for FAIRifying my fictive data set	28
9.5 ODI Data Ethics	29

1 Relevance Description

1.1 Awareness

The virtual Cybersecurity Escape Room is a great tool introducing to the relevance of being aware of cybersecurity aspects at the workspace. The following table, taken from the awareness primer of the course, describes the topics covered by the game and concrete examples from the game.

Table 2. *Virtual CySecEscape 2.0—Awareness topics and related puzzles.*

<i>Awareness topic</i>	<i>Puzzle</i>
Physical security	Unclean desk contains hints about login credentials.
Password hygiene	Easy-to-guess password.
Source code security	Source code evaluation (“hidden path”).
Information disposal	Bank account data in trash.
Securing sensitive digital data	Password reuse on sensitive file.
Public oversharing/identity theft	The missing employee is found through social media, then turns out to be impersonated.
Phishing and online banking	Phishing mail causes loss of access to bank account (game ends).

Here are some lessons learnt from the game regarding the password hygiene:

- don’t use your name as username,
- don’t use easy combinations of numbers on the keyboard as password or personal information/interests/people’s names.
- Destroy confidential information completely.
- If you use password manager don’t use the same password as used to log into the PC.
- Don’t leave the code of a drawer set on the factory default
- Social Media Oversharing: don’t share lots of personal data online

Why is it relevant?

Most part of Swiss small to medium enterprises consider that the risk of cyberattacks is low whereas a major part of them suffer from cyberattacks every year. Dozens of thousands of enterprises worldwide are hacked daily. The importance of IT systems is neglected. Most of the cybersecurity incidents are due to ignorant or careless behavior from employees or employers (Petra Maria Aspiron, 2022). Therefore, an attentive study of cyber risks and information security principles is essential to keep one safe from personal and organizational perspective.

1.2 Personal Information Security

Here are some examples of vector of risks given in the lecture (P. Asprion, P. Moriggl, 2022)

Typical Vectors of Attack (a collection...)

- **Misconfiguration**
 - Unchanged default passwords
 - Cloud credentials not set correctly
 - Unwanted components running, which compromise security model
- **Phishing**
 - Form of social engineering.
 - Try to trick target to share credentials or execute code
- **Vulnerability**
 - Weakness in a Software
 - Makes the Software exploitable
- **Weak/Compromised credentials**
 - Username/Password can be broken easily
 - Username/Password are breached by phishing, malware, same PWs etc.
 - Most common attack vector
- **Malicious Insider**
 - Insider Jobs are often hard to defend.
 - Insider steals data he has access to
 - Insider convince people to give him more data
 - Insider deploys malware (lateral movement is easy than from outside)
- **Missing or poor encryption**
 - Sensible data are transferred without encryption
 - Password can be stolen, Data can be breached
- **SQL Injection**
 - Inject a customized SQL query into a database
 - Leads to information leak
 - Can be prevented by prepared statements
- **Trojans**
 - Software pretending to be useful but being malicious
- **Cross-site Scripting (XSS)**
 - Injecting malicious code into a website
 - Impact other users view on the website

To defend one self from these vector of attacks, here are some solutions suggested in the lecture :

- Create **encrypted** message using PGP tool (<https://pgptool.org/>)

To send files in a safe and secure manner, we can encrypt them as a message using our private key and share it using public key of the receiver.
- Use **Tor** browser. «Tor browser hides your IP address and browsing activity by redirecting web traffic through a series of different routers known as nodes. Because Tor hides browsing activity and blocks tracking, it's used by whistleblowers, journalists, and others who want to protect their privacy online» (Avast.com). A disadvantage of Tor is that it significantly slows down the browsing because of the numerous hops your data is relayed through. From personal experience, it was very heavy loading and half of the screen was black.
- **Source code** security: Implementing network security solutions such as **firewalls**, Virtual Private Networks (VPN), **anti-virus**, and **anti-malware software** count as basic protection. These solutions safeguard your source code from external exploits of hackers and ensure secure data sharing between employees and data sources.
- Use transparent **search engines** like DuckDuckGo that collects less data about the user.
- Install **browser extensions**: get rid of advertisements (uBlock Origin) and other things that can be used to attack you. For example, **Privacy Badger** locates trackers and blocks them.
- **Firewalls** are personal softwares installed on the end device (between the user and the internet) that protect against access from the networks. It can help reduce the surface of attack. Recommended firewall-Softwares are for example Windows defender or Malwarebytes.
- Password hygiene: use strong passwords, don't use personal information or patterns, Rather use acronyms. Manage your passwords with **password manager** browser extension like KeepPass or Bittwarden
- To fight against fishy mails, install the **dangerzone** software.

1.3 Organizational Information Security

After studying the 2010' major breaches (special case of Microsoft in 2019) and The 2015' -- major breaches in Pharma, here's the result of our discussion and research regarding possible solutions or prevention measures against cybersecurity attacks for organizations.

- Encryption
- Give the possibility to the users/clients to report potential breach -whistleblowing
- Give guidelines on creating unguessable login/username/passwords
- Implement firewalls to avoid access to the server information
- The employee should be adequately trained on security practices and protocols
- Have a regular cybersecurity external audit
- Use the Cyber Risk Index (CRI) Calculator, to be aware of the risks -- <https://go.trendmicro.com/new-web/securityintelligence/cyber-risk/calculator.html>
- Don't put trust in the cybersecurity of another organization, ensure your own.
- Apply data retention policy (destroy/encrypt data regularly)
- Ensure that all computer rooms, communication and information systems in which confidential and personal information obtained or generated from EMA systems is stored and/or handled are protected by appropriate security measures.
- Ensure that the infrastructure is adequately protected with Antivirus software and that patch management and system vulnerabilities are regularly checked (European Medicines Agency, s.d.)



Figure 2 Top risk factors across 5 key risk areas

1.4 Terminology

Here are some important terms taken from the lecture (P. Asprion, P. Moriggl, 2022)

Virus is a type of malware. It's a code or program designed to covertly enter a device. Then, when the device runs the virus's code, the virus replicates itself within a device and transmits itself to other devices. The replicating virus can slow down the device and cause it to crash. Users may also see a large number of emails being sent from their accounts that they did not send. This is the virus spreading itself to other devices.

Trojan Horse, similar to a virus, it is an application that pretends to be good, but is designed to be malicious, does only reproduce if app is executed.

Worm is an independent piece of software. It reproduces itself without interaction.

Data Breach is an incident where sensitive or confidential data were stolen. It can be an insider job or by unauthorized access or a combination of both.

Vector of Attack is a class of ways how to get into a computer system. It can involve multiple system components, so don't think about them individually. If it is passive, it gains access without affecting the system and if it is active it breaches into the system by force.

Attack Surface is the total number of attack vectors an attacker could use against a system. In general: Small Attack Surface is good security practice. Reduce your surface by closing unused ports, removing unnecessary components etc.

Firewall is a network security device that monitors incoming and outgoing network traffic and decides whether to allow or block specific traffic based on a defined set of security rules. Firewalls establish a barrier between secured and controlled internal networks that can be trusted and untrusted outside networks, such as the Internet. A firewall can be hardware, software, software-as-a service (SaaS), public cloud, or private cloud (virtual).

Router is a network device that connects different computer networks by routing packets from one network to the other. This device is usually connected to two or more different networks. When a data packet comes to a router port, the router reads the address information in the packet to determine out which port the packet will be sent (e.g. a router provides you with the internet access by connecting your LAN with the Internet). If two hosts from different networks want to communicate with each other, they will need a router between them.

Whistleblower is a person, often an employee, who reveals information about activity within a private or public organization that is deemed illegal, immoral, illicit, unsafe or fraudulent.

Cybersecurity is a part of information security. It is measures taken to protect a computer, computer system, information assets against unauthorized access, attack or threats.

Communications Security: Protection against a threat to the technical infrastructure of a cyber system which may lead to an alteration of its characteristics to carry out activities which were not intended by its owners, designers or users.

Operations Security: Protection against the intended corruption of procedures or workflows which will have results that were unintended by its owners, designers, or users.

Physical Security: Protection against physical threats that can influence or affect the well-being of a cyber system. Examples could be physical access to servers, insertion of malicious hardware into a network, or coercion of users or their families.

Public/National Security: Protection against a threat whose origin is from within cyberspace but may threaten either physical or cyber assets in a way which will have a political, military or strategic gain for the attacker. Examples could be 'Stuxnet' or wide-scale DOS attacks on utilities, communications financial system or other critical public or industrial infrastructures.

2 Information Security Policy

“Switzerland ranks third among most targeted European countries for cyberattacks, behind Germany and the U.K., according to Switzerland insurance company Swiss Risk and Care. It’s the seventh most targeted country in the world.” (Fierce Pharma, 2022). The company of my choice is Novartis, a pharmaceutical group in Switzerland. The Information security Policy can be found in appendix.



Classifying data-level entries:

L1: for the first level of Information security, I got inspired by the Harvard policy since the public information of most companies are similar. (Harvard Information Security, 2017). I also browsed the Novartis official website to see which information is available for public. (Novartis, s.d.) This first level is the one I could add the most examples given that I could access them directly.

L2: Inspired by Novartis’ available documents on Codes, Policies and Guidelines as well as available Data protection privacy (Novartis).

For further levels, L3-L4, as the information about these levels of security are internal sensitive, I can only imagine what this type of information security can be. Some of them were again taken from the Harvard example provided in class.

L5 : This level concerns more governmental organizations as isolation from the network is a drastic measure. I couldn’t think of a concrete example for a pharmaceutical group except those provided by the Harvard example.

2.1 Definitions of information security, terms

The definitions present in this section are taken from the class on Information Security & Cybersecurity (P. Aspöhn, P. Moriggl, 2022). One must stay critical given that the formal definitions depend on the source and on the field.

Information Security Protection against the threat of theft, deletion or alteration of stored or transmitted data within a cyber system. IS ensures that, within the enterprise, information is protected against disclosure to unauthorized users (*confidentiality*), improper modification (*integrity*) and nonaccess when required (*availability*).

Information security policy is an aggregate of directives, rules, and practices that prescribes how an organization manages, protects, and distributes information.

Information security governance: the set of **responsibilities** and practices exercised by the board and **executive management** with the goal of providing strategic direction, ensuring that objectives are achieved, ascertaining that risk is managed appropriately and verifying that the enterprise’s resources are used responsibly.

Information security program: the overall combination of technical, operational and procedural measures and management structures implemented to provide for the confidentiality, integrity and availability of information based on business requirements and risk analysis.

Information security testing tools: tools used to test the accuracy and completeness of an enterprise’s cybersecurity practices and controls. Monitoring/audit activities.



Figure 3 CIA triad

2.2 Own perspective

"Data is the new gold. Data becomes a valuable commodity and must be professionally protected - confidentiality, integrity and accessibility are now a must!" (P. Asprion, P. Moriggl, 2022).

Generally, in private or professional environments data worth protecting are personally identifiable information such as names, addresses, emails, phone numbers, health information, bank details. It can also be information that makes the institution unique and competitive on the market, such as intellectual property. This kind of data should be carefully protected as if it is managed incorrectly can cause important social, reputational, financial, physical or material harm to enterprises or individuals.

2.3 Organizational perspective

Data worthy of protection is classified according to government and institutional requirements.

"Under FDA regulations, an Institutional Review Board (IRB) is group that has been formally designated to review and monitor biomedical research involving human subjects" (FDA , s.d.)

Here's a list of some influential organizations that work on requirements for data handling and protection:

- **ENISA** - The European Union Agency for Cybersecurity contributes to EU cyber policy, cyber and digital security.
- **NIST** - National Institute of Standards and Technology), provides policies, laws and regulations on Information Security
- **ISO** - International Organization for Standardization is an independent, non-governmental international organization
- **ISACA** – Information Systems Audit and Control Association reunites experts in disciplines of IS/IT audit, risk, security and governance as well as educators, consultants and regulators.

How data protection measures are implemented at Novartis ? We can read about it in their Privacy Policy:
“We have implemented appropriate technical and organizational measures designed to provide an adequate level of security and confidentiality to your Personal Information. The purpose of these measures is to protect Personal Information against accidental or unlawful destruction or alteration, accidental loss, unauthorized disclosure or access and against other unlawful forms of processing.” (Novartis, 2021)

2.4 Additional thoughts

In the post-COVID era, one of the factors to consider about data protection is “home office”. Regarding pharmaceutical industries, “any pharmaceutical processes that require labs, such as clinical research or toxicology, are not viable for remote work due to security and ethical concerns surrounding animal and human trials. Nearly all other pharmaceutical processes can be executed remotely.” (Pharma Manufacturing Magazine, 2021)

For tasks that are compatible with remote working, several strategies can be implemented by the companies to keep their data safe, e.g. installing VPN, institute end-to-end encryption, provide employees training, develop cybersecurity policy and more (some examples are discussed in the appendix in the section Coaching session 2: Cybersecurity).

3 Data Management Plan

A data management plan, or **DMP**, is a formal document that outlines what one does with data during and after a research project. Many funding agencies, especially government funding sources, require a DMP as part of their application processes.

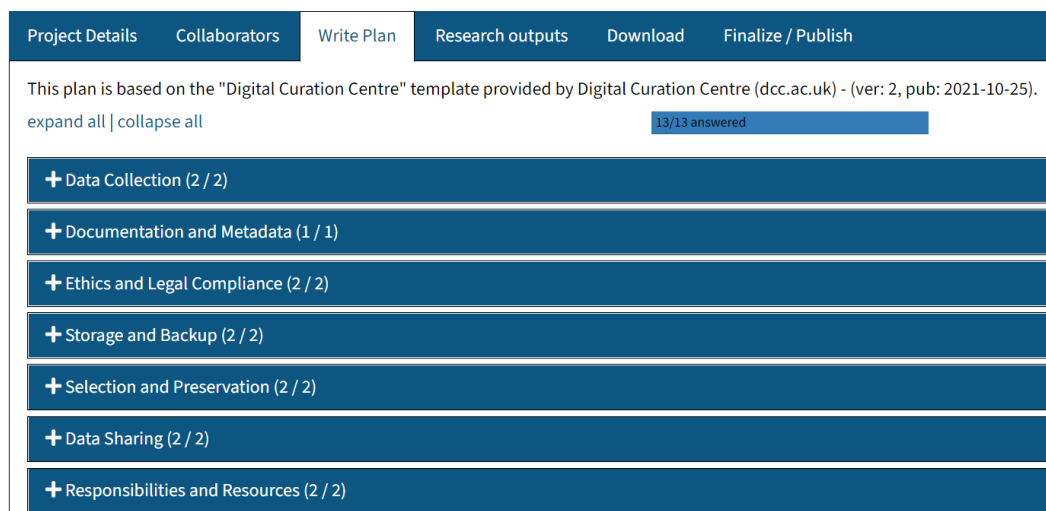
The following section and the DMP in the appendix result from a groupwork with Siddiqui Ali using the DMP Tool at <https://dmptool.org/>.

3.1 Project scenario

Our project scenario is a research project at ZHAW in collaboration with the University Hospital Zürich. In our master projects, we must collect human physiological measurements to track health conditions and use this data for creation of models for prediction of certain diseases. As we handle an important amount of individuals' data, we need to create a Data Management Plan for this project.

The motivation for this research project is treating a particular neurological condition (delirium) in USZ patients. Collecting physiological measurements like heartbeat, transpiration, brain activity before a crisis can help us create a model and be able to predict components that cause patients to suffer from this condition. In order to collect data, we doctors that treat these patients take measurements at the hospital and then transfer anonymized data to us (Applied Computational Life Science researchers). Modelling and prediction for this project will allow better prevention and treatment for patients suffering from the condition.

Physiological data collection for research at the USZ



Project Details Collaborators **Write Plan** Research outputs Download Finalize / Publish

This plan is based on the "Digital Curation Centre" template provided by Digital Curation Centre (dcc.ac.uk) - (ver: 2, pub: 2021-10-25).
expand all | collapse all 13/13 answered

- + Data Collection (2 / 2)
- + Documentation and Metadata (1 / 1)
- + Ethics and Legal Compliance (2 / 2)
- + Storage and Backup (2 / 2)
- + Selection and Preservation (2 / 2)
- + Data Sharing (2 / 2)
- + Responsibilities and Resources (2 / 2)

Figure 4 Outline of topics covered in a DMP

On Figure 5, one can see different drivers that can lead us to decisions in our DMP. In our project, the main drivers are Privacy, Compliance, and Information Security, as we must deal with personal, confidential and sensitive health records from patients. A DMP also helps us organize data collection during experiments so one can use this data effectively for research purposes.

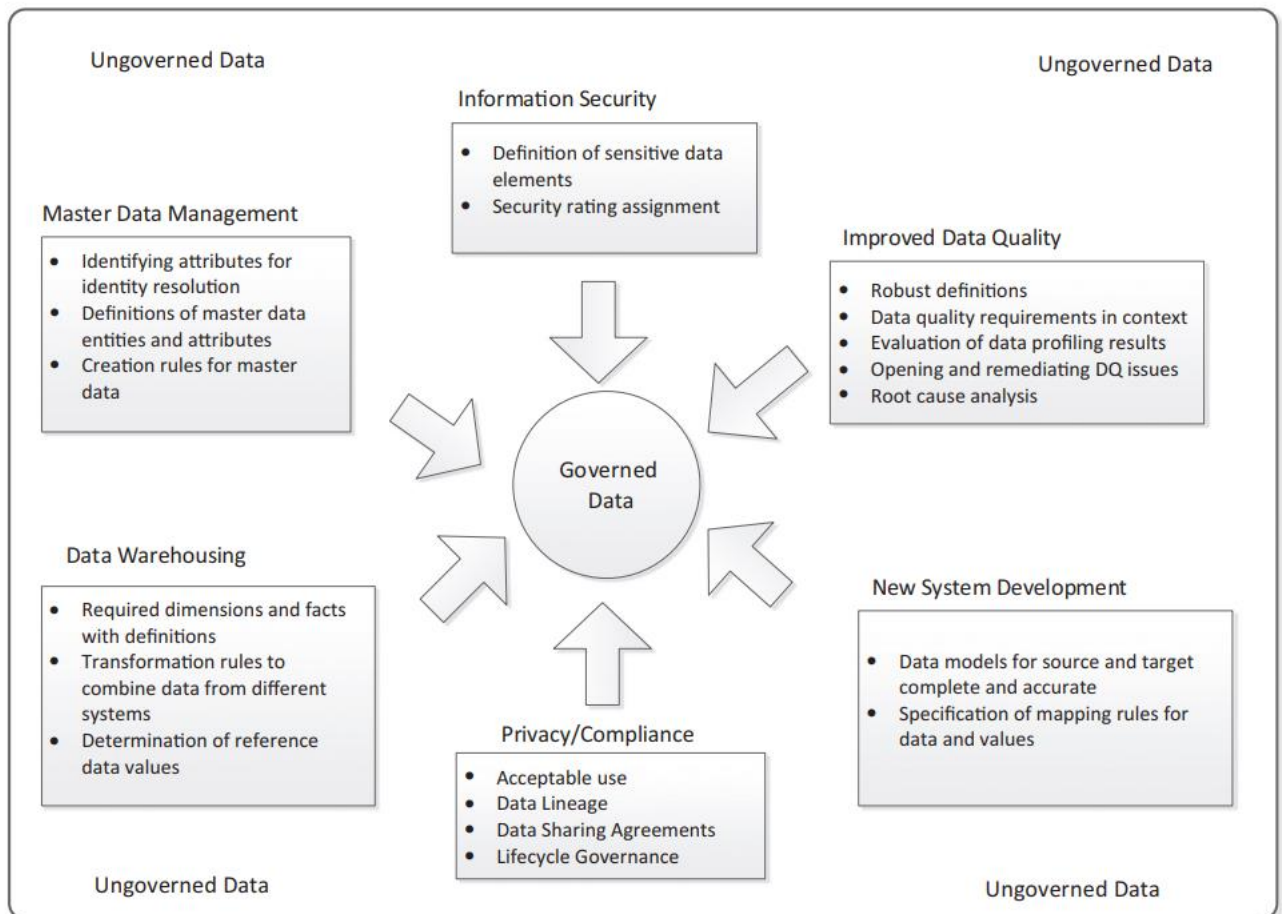


Figure 5 Drivers for moving data from an ungoverned state to a governed state

3.2 Data Perspective and DMP description

Data Collection: the types of data we collect includes quantitative and qualitative data extracted from studies conducted at the hospital. Data is collected in different formats (e.g. excel, txt, jpeg files) and named using a standardized convention.

Documentation and Metadata: data should always be accompanied by a readme file with standards and instructions on how to read the data and how to reproduce the experiments.

Storage and Backup: all data is uploaded to a cloud-server as well as on local drives. All files should be automatically backed up on a regular basis.

Ethics and Legal Compliance: consent is required from participants before storing data as a consent questionnaire describing all the legal implications. Personal data is anonymized. Intellectual Property rights are owned by the researches that conduct experiences and the USZ.

Selection and Preservation: some data is defined as of long-term relevance and should be preserved longer than other type of data (e.g. 10 years). Observations, findings and results will also be kept for Long-term, whereas processed data for example will be kept for 5 years only. This goal is met with archiving the data.

Data Sharing: digital search engine will allow university researchers to access the data for research purposes under conditions of confidentiality. For sharing data the following aspects have to be considered: credits and citations, requests handling as data request form, reports from unauthorized attempts, restrictions on data sharing and data sharing agreements with third parties.

Responsibilities and Resources: directors of departments, principal investigators of each experiment and Data Stewards are responsible for implementing the DMP. Data Management Sharing Plan includes metadata about regulations required to handle and share data according to the DMP.

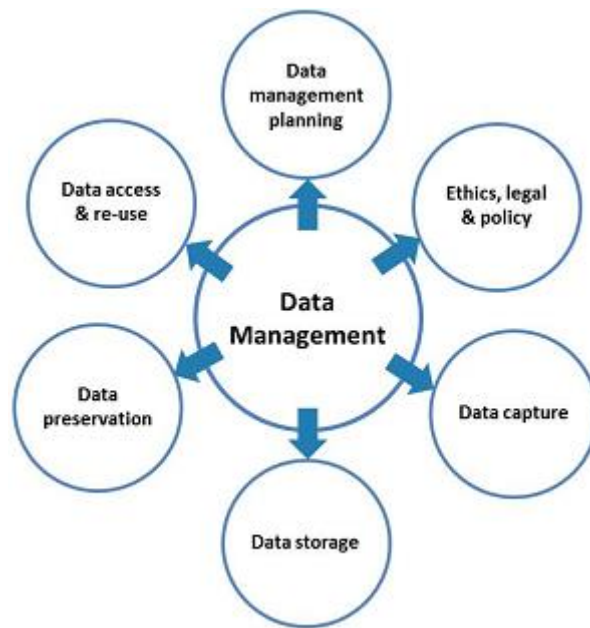


Figure 6 Data Management Plan

4 Fair Guiding Principles

Let us first examine what it means for data to be FAIR. (Wilkinson, 2016)

To be **Findable**:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

To be **Accessible**:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

To be **Interoperable**:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To be **Re-usable**:

- R1. (meta)data have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

4.1 Organizing data

Exploring methods to organize data at a data level, one can start with creating the following **folder structure**.

One way of keeping datasets easily findable and comprehensible is to keep them at the correct folders with meaningful names as in the example: the outer folder should be a location where the data was collected (e.g. Hospital, research lab) and further layers are data itself and the analysis of the data. To further ease the lecture of files inside the folders, use file naming conventions.

Example of a file naming convention:

YYYYMMDD_Researcher_name_type_nb_of_patient

20230308_GOLUBEVA_heartbeat_1384359

YYYYMMDD_Research_name_date_nb_of_patient_version

2023_GOLUBEVA_1384359_original

2023_GOLUBEVA_1384359_processedv1

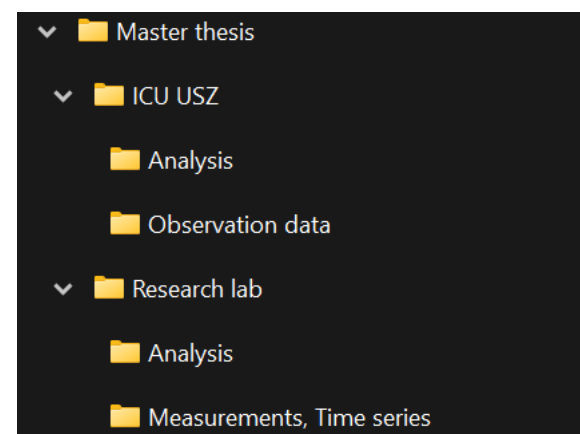


Figure 7 Folder structure

4.2 Metadata structure

After creating a dummy dataset for possible output of my project, I separated descriptive, administrative, and structural metadata by color code:

Project name	Researcher	Date	Time	Format	Type	Description	Nb_participant	Name
Delirium study	GOLUBEVA	2023.03.08	11:14 AM	Image	T1	Hippocampus	13452749	Anonymized
Delirium study	GOLUBEVA	2023.03.09	12:14 PM	Image	T2	Amygdala	13452750	Anonymized
Delirium study	GOLUBEVA	2023.03.10	1:14 PM	Time series	heartbeat	original	13452751	Anonymized
Delirium study	GOLUBEVA	2023.03.11	2:14 PM	Time series	transpiration	adapted	13452752	Anonymized

Color code	Metadata Type
Red	Structural metadata
Yellow	Descriptive metadata
Grey	Administrative metadata

Figure 8 Types of metadata: example

The following definitions of different metadata types were taken from the lecture (P. Aspiron, P. Moriggl, 2022).

Administrative metadata are data about a project or resource that are relevant for managing it; for example, project/ resource owner, principal investigator, project collaborators, funder, project period, etc. They are usually assigned to the data, before you collect or create them.

Descriptive or citation metadata are data about a dataset or resource that allow people to discover and identify it; for example, authors, title, abstract, keywords, persistent identifier, related publications, etc. Structural metadata are data about how a dataset or resource came about, but also how it is internally structured. Structural metadata describe, for example, the unit of analysis, collection method, sampling procedure, sample size, categories, variables, etc.

Structural metadata have to be gathered by the researchers according to best practice in their research community and will be published together with the data. Descriptive and structural metadata should be added continuously throughout the project.

4.3 Codebook for example data set

Besides a data map (Figure 9) one can create a codebook that describes variables used in your dataset to facilitate data reading and make it FAIR, in particular accessible.

To write the codebook for my example I used a [template](#) available online. In the data map, one can see the relationships between variables in my data set.

The data set is presented in Figure 8 and the codebook can be found in Appendix.

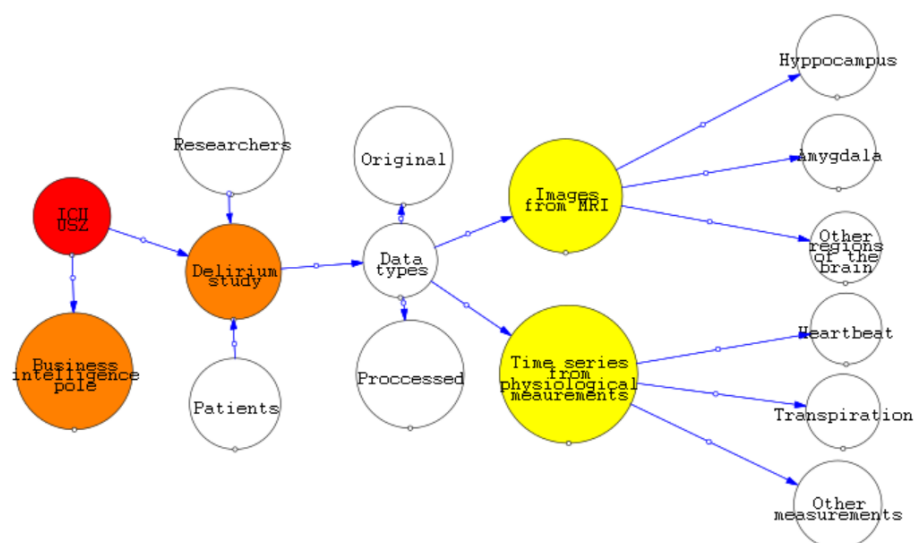


Figure 9 Data map

4.4 FAIR elements for example dataset

How can FAIR principles be applied to this example data set?

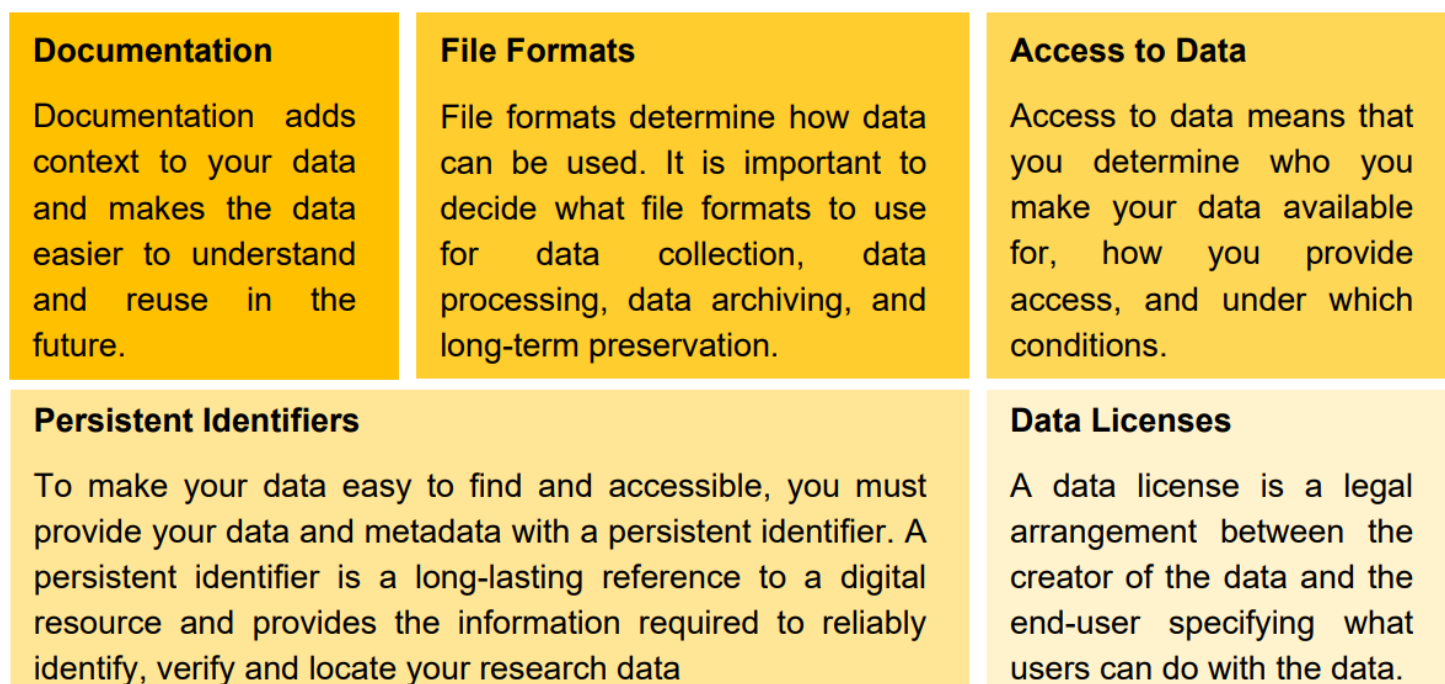


Figure 10 How to FAIR

Here are some examples of actions that can be taken to store and share my data set according to FAIR principles or to FAIRify data:

- Document data (repositories) : data level documentation, data map, codebook
- Choosing appropriate file formats (Figure 11)
- Adding metadata(administrative, descriptive and structural), readme file
- Giving access to data (e.g. figshare, zenodo, GitLAB)
- Licensing the data or adding a persistent identifier(DOI)

Containers:	TAR, GZIP, ZIP
Databases:	XML, CSV, JSON
Geospatial:	SHP, DBF, GeoTIFF, NetCDF
Video:	MPEG, AVI, MXF, MKV
Sounds:	WAVE, AIFF, MP3, MXF, FLAC
Statistics:	DTA, POR, SAS, SAV
Images:	TIFF, JPEG 2000, PDF, PNG, GIF, BMP, SVG
Tabular data:	CSV, TXT
Text:	XML, PDF/A, HTML, JSON, TXT, RTF
Web archive:	WARC

Figure 11 Example of preferred FAIR file formats

5 Data Ethics Canvas

5.1 Ethics terminology

Offered in the lecture, here are some useful terms to enter the matter of ethics which are relevant for our study case.

What is data ethics?

In a paper for the Royal Society in late 2016, researchers Luciano Floridi and Mariarosaria Toddeo define data ethics as:

*"The branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)."*⁴

At the ODI, we believe short and accessible definitions are necessary to include more people in debates that impact them. We suggest and use a different data ethics definition of: **a branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use.**

Figure 12 Data ethics

Information Ethics -the subject of information ethics is the morality of those who offer and use information and communication technology (ICT), application systems and new media. It inquires how these persons, groups and organizations behave in aspects of morality and how they should behave. According to Rafael Capurro, we can divide the discipline into computer ethics, net ethics, and media ethics (the lecturer prefers the term of new media ethics).

Business Ethics - branch of applied ethics that studies the moral dimensions of commercial activity, frequently but not exclusively with respect to corporations. It encompasses an extremely broad range of issues, including whether and how corporations – as distinct from their officers or shareholders – are moral agents.

Animal Ethics deals with the moral duties of humans towards animals and with the moral rights of animals. The ability to suffer is an important moral and ethical argument. It can be used to justify species-appropriate animal farming or a ban of animal farming and animal use.

Machine Ethics refers to the morality of semi-autonomous or autonomous machines, the morality of certain robots or bots is one example. Hence these machines are moral agents. They decide and act in situations where they are left to their own devices, either by following pre-defined rules or by comparing the case to selected case models, or as machines capable of learning and deriving rules. Moral machines have been known for some years, at least as prototypes.

Technology Ethics relates to moral issues of the use of technique and technology. It can focus on the technology of vehicles or weapons as well as on nanotechnology. There are manifold relations to science ethics. In the information society, technology ethics is also closely connected to information ethics.

5.2 Ethics Review

The main goals of ethical considerations are building trust and challenging common assumptions. Today there are many existing data ethics frameworks.

The Institutional Review Board (IRB) is an administrative body established to protect the rights and welfare of human research subjects recruited to participate in research activities conducted under the auspices of the institution with which it is affiliated.

Criteria for IRB Approval of a Human Research Study (ULCA, 2021) :

To be approved, a study must meet the criteria listed below (some examples):

1. Risks to subjects are minimized.
2. Risks to subjects are reasonable in relation to anticipated benefits
3. Selection of subjects is equitable.
4. Informed consent will be sought or waived in accordance with 45 CFR 46.116—and 21 CFR 50.25 for FDA-regulated research
5. Informed consent will be documented, or documentation waived in accordance with 45 CFR 46.117—and 21 CFR 50.27 for FDA-regulated research
6. Provisions for monitoring collected data are adequate to ensure the safety of subjects.
7. Provisions to protect privacy of subjects are adequate. Provisions to maintain confidentiality of data are adequate.

5.3 Reflection on Data Ethics Canvas

The Data Ethics Canvas is in the Appendix.

Here's a summary of the ethical overview of my project:

Data sources Rights around data sources Limitations in data sources Ethical and legislative context	The ethical data framework is given by the legislation in place at the Hospital (e.g. Hippocratic Oath) and at ZHAW
Your reasons for using data Positive effects on people Negative effects on people Minimizing negative Impact	The purpose of the project is to improve public health and develop medication and therapies. Negative impact is minimized by anonymity and fundamental moral and ethical principles of a doctor's behavior
Engaging with people Communicating your purpose Openness and Transparency Sharing data with others	As much as we want to make the project visible and transparent, we want to save patients' personal sensitive data. Essential aspects are covered in the consent form that patients must sign if they want to partake in the project.
Ongoing implementation Reviews and iterations Your Actions	The plan of actions is especially available for the grants/funds organizations. It will not be publicly published before getting funding. The ethical considerations are updated and discussed during ethics committees and based on feedback from patients and researchers.

6 Conclusion and Reflection

6.1 Content

As soon as data is gathered, numerous underlying factors become crucial, including effective data management, user and provider rights and obligations, security, usage ethics, and more. As data volume has greatly increased over the past few decades, handling and protection have become even more important. It is even more relevant when personal or health information is gathered from individuals, as in the pharmaceutical or medical industries, for instance.

In terms of personal security, one should recognize that, as cybercrimes emerge, one is responsible for reducing the surface of attack at home. Best practices should be implemented at the organizational level too to protect the company, its employees, and clients from cyberattacks, breaches, and threats originating from third parties or caused by simple negligence. Every company is encouraged to comply to regulations such as Information Security policy to keep sensitive data secure.

Data is the new gold in both the business world and the academic world. It must be managed in accordance with established rules, such as the FAIR principles, in order to be handled responsibly. Data management's increasing relevance has even given rise to new positions like Data Steward.


When working with data, ethics is always a core consideration. Laws are properly implemented in accordance with moral guidelines for data exchange, handling, and usage thanks to ethical evaluations, frameworks, and institutions. Each entrepreneur, project manager, principal investigator, and researcher is likewise accountable for understanding the ethical consequences of their activity.

6.2 Coaching Sessions

Coaching sessions were a fantastic setting for hands-on learning and practical application of all the lecture-learned principles. I gained a new perspective on all the factors that businesses must take into account from various points of view, like data level management, ethics, legal requirements, policies, and legislation, through practical projects like establishing a DMP, an ISP, or a Data Ethics Canvas. I was able to place myself in the role of a project manager and see all the assumptions I have to address that were left implicit for me prior to this experience.

6.3 Personal view

Doing my Masters in Applied Computational Life Sciences and aiming to pursue my career in research I am going to deal with human and animal data. Thanks to this course, I now have a valuable skill set and a perspective that will help me approach data-related problems with more caution and thought. Working with sensitive data has ethical, legal, and organizational repercussions, which I am now aware of. The course material helped me to understand the importance of cybersecurity in both my personal and professional life.



Data is the oil, some say the gold, of the 21st Century – the raw material that our economies, societies and democracies are increasingly being built on.

JOE KAESER, CEO, SIEMENS

7 Bibliography

- Avast.com. (kein Datum). *The Dark Web Browser: What Is Tor, Is It Safe, and How to Use It*. Von <https://www.avast.com/c-tor-dark-web-browser#:~:text=The%20Tor%20Browser%20hides%20your,to%20protect%20their%20privacy%20online.>
- European Comission. (kein Datum). *Data management*. Von https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template
- European Medicines Agency. (kein Datum). *EudraVigilance: security principles and responsibilities*. Von <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance/eudravigilance-security-principles-responsibilities>
- FDA. (2019). *Institutional Review Boards (IRBs) and Protection of Human Subjects in Clinical Trials*. Von <https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/institutional-review-boards-irbs-and-protection-human-subjects-clinical-trials>
- Fierce Pharma. (2022). Von <https://www.fiercepharma.com/pharma/novartis-hit-cyber-attack-says-no-sensitive-data-was-compromised>
- Force11. (2020). *The FAIR Data Principles*. Von <https://force11.org/info/the-fair-data-principles/>
- Harward Information Security. (2017). *Information Security Quick Reference Guide*. Von <https://mslscommunitycentre.ch/pluginfile.php/19842/course/section/2063/HarvardPolicy.pdf>
- Identity Theft Resource Center. (2019). *Microsoft Email Breach and the Need for Password Security*. Von <https://www.idtheftcenter.org/post/microsoft-email-breach-and-the-need-for-password-security/>
- LifeLock by norton. (2021). Von <https://lifelock.norton.com/learn/data-breaches/microsoft-exposed-250-million-customer-records>
- Novartis. (kein Datum). Von <https://www.novartis.com/>
- Novartis . (kein Datum). *Data privacy*. Von <https://www.pensionskassen-novartis.ch/en/pension-funds/data-privacy>
- Novartis. (2021). *Privacy Policy*. Von <https://www.novartis.com/privacy/privacy-policy>
- Novartis. (kein Datum). *Codes, Policies and Guidelines*. Von <https://www.novartis.com/esg/reporting/codes-policies-and-guidelines>
- Novartis. (kein Datum). *Minimum Information Security Controls for Third Parties*. Von https://www.novartis.com/sites/novartis_com/files/novartis-minimum-information-security-controls.pdf
- P. Asprion, P. Moriggl. (2022). *Lecture Slides, D4 Data and Ethics*.
- Pharma Manufacturing Magazine. (2021). Von <https://www.pharmamanufacturing.com/information-technology/security/article/11291686/making-work-from-home-work-in-pharma>
- ULCA. (2021). *Requirements for IRB Review and Approval*. Von https://ora.research.ucla.edu/OHRPP/Documents/Policy/12/IRB_Review_Requirements.pdf
- Wilkinson, M. D. (2016). *The FAIR Guiding Principles for*. Von <https://www.nature.com/articles/sdata201618>

8 List of Figures

Figure 1 – Cybersecurity, source : <https://losspreventionmedia.com/an-a-z-plan-for-corporate-cybersecurity-success/>

Figure 2 - Top risk factors across 5 key risk areas, source : https://www.trendmicro.com/en_us/security-intelligence/breaking-news/cyber-risk-index.html

Figure 3 – What is the CIA triad ? <https://www.varonis.com/blog/cia-triad>

Figure 4 - Data Management Plan : <https://dmp.qut.edu.au/faq>

Figure 5 – Drivers for moving data from an ungoverned state to a governed state, from P. Asprion, P. Moriggl. (2022). *Lecture Slides, D4 Data and Ethics*.

Figure 6 – Data Management Plan <https://dmp.qut.edu.au/faq>

Figure 7 – Folder structure, personal picture

Figure 8 - Types of metadata: example, personal picture

Figure 9 – Data map, personal picture created with Vensim, visualization software

Figure 10 – How to FAIR, from P. Asprion, P. Moriggl. (2022). *Lecture Slides, D4 Data and Ethics*.

Figure 11 - Example of preferred FAIR file formats, from P. Asprion, P. Moriggl. (2022). *Lecture Slides, D4 Data and Ethics*.

Figure 12 Data ethics <http://rsta.royalsocietypublishing.org/content/374/2083/20160360#sec-1>

9 Appendix

9.1 Information Security Policy

CLASSIFICATION				
L1 Information intended and released for public use.	L2 Information that may be shared only within the project community.	L3 Confidential and sensitive information, intended only for those with a “business need to know.”	L4 High-risk information that requires strict controls	L5 Extremely sensitive information requiring specific controls and isolation from the network..
The company intentionally provides this information to the public.	The company chooses to keep this information private, but its disclosure would not cause material harm.	Disclosure of this information beyond intended recipients might cause material harm to individuals or the company.	Disclosure of this information beyond specified recipients would likely cause serious harm to individuals or the company.	Disclosure of this information could cause criminal liability; loss of insurability or employability; or severe social, psychological, reputational, financial, or other harm to an individual or group.
<ul style="list-style-type: none"> - Published articles in Live Magazine/Press releases - Currently offered treatments/products - Division information, organizational structure - List of collaborators - Published annual reports - Currently recruiting clinical trials - Commercial model, values - Statistics of employees and collaborators 	<ul style="list-style-type: none"> - Department policies and procedures - Drafts of research papers - Pre-release articles - Patent applications - Grant applications - Work papers - Non-public building plans or layouts (excluding L3 or L4 items) 	<ul style="list-style-type: none"> - Personnel records - Intranet website/portal - Internal guidelines - Emails - Donor information - Non-public legal work - Budget /financial transactions information 	<ul style="list-style-type: none"> - Research data - Research projects ideas - Personal data from individuals - Surveys results - Passwords and PINs - System credentials - Private encryption keys - Individually identifiable health or medical information - Individually identifiable research data - Security system procedures 	<ul style="list-style-type: none"> - Research data classified as Level 5 by the IRB - Information or research under a contract stipulating specific security controls beyond L4

9.2 Coaching session 2: Cybersecurity

McAfee recommends the following best practices:

- Do not allow RDP connections over the open internet.
- Use complex passwords as well as multi-factor authentication.
- Lock out users and block or timeout IPs that have too many failed login attempts.
- Use an RDP gateway.
- Limit domain admin account access.
- Minimize the number of local admins.
- Use a firewall to restrict access.
- Enable restricted admin mode.
- Enable Network Level Authentication (NLA).
- Ensure that local administrator accounts are unique and restrict the users who can login using RDP.
- Consider placement within the network.
- Consider using an account-naming convention that does not reveal organizational information.
- Target Data Breach, How Target Almost Lost Everything
- What to Do to Protect Yourself When Buying from Retail Stores
- Use only one credit card for retail purchases and monitor your statements carefully each month.
- Review bank statements and your credit report regularly to scan for fraudulent activity.
- Invest in credit monitoring and consider a credit freeze where new accounts cannot be opened without your permission.
- Keep all your devices updated with antivirus software and run scans often.
- Use common sense and watch for suspicious scam emails that push you to click a link or download an attachment.

9.3 Data Management Plan

Physiological data collection for research at the USZ

Data Collection

What data will you collect or create?

We will collect physiological data from experiment participants such as

- ECG signal from the brain, heartbeat, temperature, blood pressure, transpiration rate - continuous signal
- age, body mass, height - numeric data type
- scope of physical activity, gender - factor data type
- images from MRI scanner - jpeg files

All are stored as xlsx, csv files, on a hard drive and in an Electronic Lab Notebook. This allows data to be stored, backed up, shared and long-term accessed. The files are sent to the Hospital's cloud with set deadlines of expiration of data access.

How will the data be collected or created?

Participants give informed consent before participating. For personal information (age, gender, physical activity) participants fill in a questionnaire prior to the experiments.

Then measures are taken from the participants using thermometer, electrocardiograph, manometer, eye tracker, neuroimaging techniques (MRI scanner, EEG or PET) and other physiological measurement tools. As participants are regularly examined by doctors, they also provide their blood and brain drug concentration over the period of experiments. Each participant has an assigned number and folder in the data set. We will use ISO standards for dates and a naming convention :

File Naming Convention

- Project lead's last name or initials.
- File creator's last name or initials.
- Project name/acronym.
- Date file created/generated (in YYYY-MM-DD format)
- Version number

Documentation and Metadata

What documentation and metadata will accompany the data?

An excel file with participants' names and associated numbers will accompany the data files to decode participants and their attributes. A readme text file with the following information will also accompany the study :

- statistics about the data (total number of observations)
- models of the tools used type of data
- instructions for researchers.

An addition Word file will include the questionnaires templates with predefined questions.

Ethics and Legal Compliance

How will you manage any ethical issues?

The consent is required from participants before storing data as a consent questionnaire describing all the legal implications. Anonimization can be done by hashing/encryption of personal and identifiable data.

- the basic database's (with participants' names) access is regulated for patient's doctors.

- another data without identification (names, date of birth) can be accessed with a barcode or a random number assigned to a patient.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The data is own by the USZ and Intellectual Property rights are owned by the researches that conduct experiences and USZ as a whole.

Storage and Backup

How will the data be stored and backed up during the research?

- The Data will be stored on the Cloud-server of the at the research organization and hospital (project collaboration platform).
- If the data is processed on local drives, the files and folders need to be saved in the following folder: `:/projects/data`
- Set up of the data space in the Project Collaborative Platform
 - o Implementation of the UUID generator
 - o DOI registration request
 - o Preparation of templates for:
 - o Data descriptor (general, text format, pdf output)
 - o Metadata: text template, spreadsheet template
- This folder is secured and all files are automatically uploaded and backed-up.
- It will require 500gigabyte, hence we will require additional storage.
- The data will be backed up end of the week, and uploaded automatically to the server.
- The IT admin is responsible and recovery of the data.
- Barring any major incidents on the server, the data can be retrieved from the cloud-server if any mishaps occur on the local data.

How will you manage access and security?

- Only authorized user will be given access to the data i.e.
 - o the project supervisor
 - o the project leader
 - o the project collaborators/ hospitals
- A password protected will be provided to the authorized person
 - two-factor security of access will be established using the Authenticator app
- Collaborators will need to provide an email address and a phone number for accessing the data
- Data encryption will be put in place before the data is uploaded to the main secure system
- GDPR will be applied and personal data
- Files uploaded and backed-up to the secure server will appear in a color-coded format (e.g. green-> for uploaded and backed up, blue-> for only local files, red-> unsaved files)

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

- Patient critical Meta data and business critical data will be long-term data on:
 - o Long-term is defined to 10 years

- Observations, findings and Results will also be kept for Long-term
- Each dataset is initially assigned to a unique ID and classification, automatically generated through a Universally Unique Identifier
- (UUID) application.
- Processed data will be kept for 5 years
- The data can be used for future research studies
- Data can be used Research Paper publications in international journals
- No additional cost is associated with the maintenance of data on the main server (10 years)
- A dedicated software will be deployed for saving the data with specific identifiers which will allow the definitions of the data retain-ability
- A training will be established for familiarizing the project participant on different classes of the data

What is the long-term preservation plan for the dataset?

- Once the project is completed, it will be the responsibility of the project leader to close the project.
- When the project is closed, the project supervisor will sign-off the main server location to be moved to the archived repository
- Only authorized user will be given the access to the archive folder
- No additional cost will be accrued for the data preservation
- It is the responsibility of the project leader to prepare data for sharing / preservation
- Internal audits of the folder may take place each quarter for the assuring the quality of the data and folders

Data Sharing

How will you share the data?

In general, only data related to publications will be made openly available In general, the project-leader will decide on a case-by-case basis which

data can be released to avoid issues related to IP rights protection or access.

- Digital search engine will allow university researchers to look for keywords related to the projects which will allow them to find out the existence of the data.
- The data could be shared within the ZHAW researchers, under the condition of maintaining the confidentiality and not divulging the trade secrets the data can only be used for research and non-commercial purposes.
- Credits and citation will be given by the future users to the project team
- A handle request will be established and access to only the relevant files will provided
- The data-request form will be established for access the data in future
- The data-log will be established, and password secured will be put in place to identify who is accessing the data and when it is access
- Reports from unauthorized attempts to access the data will also be logged-in
- An NDA will be established so that the data is not dispersed to any third-parties

Are any restrictions on data sharing required?

- Personal and confidential information and data will be restricted before sharing.

- Normalization of any identifiers will be established during the project i.e. keeping the identity of any individual secured. Specific agreements with the Editors of scientific/technological journals will be considered and provided.
- Results of the study and any improvements will only be exclusive between ZHAW and the external partner
 - o Hospital
 - o Machine manufacturers
 - o Software manufacturers
- A data sharing agreement will be required if the request is coming from outside of ZHAW
- NDA will still be established between external parties and the project leader

Responsibilities and Resources

Who will be responsible for data management?

The directors of departments, principal investigators of each experiment and Data Stewards are responsible for implementing the DMP, and

ensuring it is reviewed and revised. Researchers are trained by their superiors in order to collect, store and handle data in accordance with the DMP.

Data ownership and responsibilities for research data management will be part of an agreement between the research organization and the hospital.

What resources will you require to deliver your plan?

A data steward is required to set the right training for the researchers. DMSP resources will include

- Onboarding and Offboarding Checklists
- ELN Resources (Electronic Lab Notebook)
- Metadata Guide

9.4 Codebook for FAIRifying my fictive data set

Project Description

In this project we collect physiological measurements from patients suffering from delirium in order to study, model, predict and prevent crisis.

Study design and data processing

Collection of the raw data

Data is collected through an MRI scanner that creates T1 type of images and a Biopac software that produces time series of heartbeat and transpiration rate.

Creating the tidy datafile

1. download the data
2. denoise data
3. structure data such that each observation is at its own row, each variable at its own column and each value in its own cell.

Description of the variables in the data

There are 8 variables including project name, researcher name, date, time, format, type and description of data collected, identification number of the participant and name. There are to this date 4 observations (in the example data set)

For further analysis and data description one can use `mean()`, `sd()`, `summary()` commands in R for statistics.

Variable Format (repeat this section for all variables in the dataset)

Variable describes the format in which the data is saved (image or timeseries).

Some information on the variable:

- Class of the variable : string/character
- Unique levels of the variable : image, time series

Variable Description (repeat this section for all variables in the dataset)

Variable describes which region of the brain is in the image or which version of the time series is saved.

Some information on the variable:

- Class of the variable : string
- Unique values/levels of the variable : brain regions/ version of time series file

9.5 ODI Data Ethics

