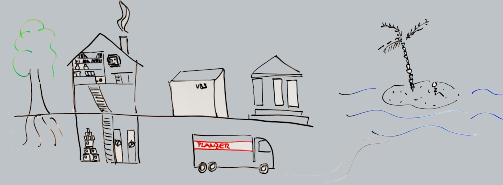


A short story about Data

DANIEL DATA is a master student. He has to deal with lots of different data every day, even if he's not always aware. Some has to do with his research project, other data has to do with his rent, his friends, his life. Let's follow Daniel and see how he organizes himself. The same principles help you to organize research data.



Active Data Management

Daniel rents a room in a shared flat. Documents he is working on and correspondence that he has to answer promptly are scattered on his desk. Other work documents are stacked in a tray. His bank and insurance documents are in binders on a shelf.

Cache / interim Archive

The shelf is now full. To make space, Daniel packs up the folders from his undergraduate studies and moves them to the basement. He seldom uses them, but he can't dispose of them yet. He also moves two folders of insurance documents, which he does not need but is obliged to keep for 5 years. Use of the basement is not for free, but it offers enough space. The documents are still available, even though it takes a little longer to reach them.

Shared memory

Daniel is the main tenant of the WG. He receives all the bills. In order for his roommates to be able to see them at any time, Daniel has stored a folder in the living room, in which copies of these documents are filed. His roommates can also file copies of expenses that they have made for the WG in this folder. At the end of the year they will settle the difference.

Security / Sensitive Data

The basement is closed by a simple door and the flat-mates also have keys; Daniel cannot be sure that his documents are not viewed by third parties. He does not want to deposit his bank folders here long-term. He also does not want to keep here the inheritance of his grandfather – a successful writer: he has very personal letters to friends, some of whom are still alive. Even worse, the basement is not fire-proof! in case of a fire everything would be lost!

Data curating / Archiving

Many valuable family documents and inherited art cannot be stored in the basement because the air is too humid. Daniel finds out that he can store his grandfather's inheritance, photo albums and family Bible in the library for a small fee. Here, all items are neatly catalogued, packed in an acid-free box and stored in a dry room. For the art he finds a museum that can keep it safe in optimal conditions. In return, the museum may exhibit it.

Delete / Share

Daniel has completed his studies and will start a PhD. He will move soon. The new apartment is larger, but doesn't have a spacious basement. Time to clean up! He no longer needs the documents from his first semesters; he disposes of them. Likewise, he disposes of all insurance documents older than 5 years. Daniel finds a few unopened boxes from his previous move, filled with old DVDs and folders from his schooldays. Daniel cannot use all this anymore and disposes of the boxes as they are. He also finds books from the undergraduate studies; he can pass these on to his younger cousin, who just started studying on the same subject.





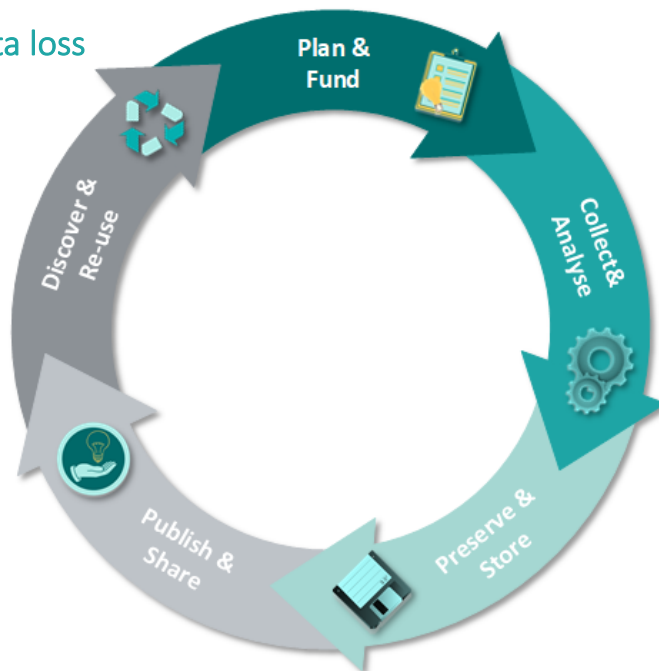
Data Management Plan



The Data Management Plan (DMP) is the roadmap to all the data that will be generated and used during your research.

It is a **tool** to help you prepare for data management in the research process. It's also a **checklist** to remind you of important aspects of handling your research data.

- ✓ Manage data effectively
- ✓ Reduce the risk of data loss
- ✓ Ensure data usability



✓ Nearly all funding agencies require a DMP as an integral part of the grant proposal

Advantages

By starting early in your project with data management planning:

- you can assign roles and responsibilities for data management in your team
- you can estimate necessary infrastructure, services, legal advice, and budget costs (sometimes reimbursable by your funder)
- you increase research efficiency, saving time and resources in the long run
- you prevent loss of data by enhancing data security
- you avoid data duplication through well-arranged data organization
- with good data organization and documentation you ensure continuity when things change and project members find the data they need more easily
- you can increase the visibility and the outreach of your research data through data publication
- you enhance the reproducibility and verifiability of your research



Data Organization



Organize your data to **FIND, ACCESS, USE and REPRODUCE** it!*

Choose a logical and consistent file organization from the beginning, one that allows you and others to **easily locate**, **access** and **use** your data, to **avoid duplication**, and ensure that your data can be **backed up**.

*According to FAIR principles



USE A FOLDER STRUCTURE

- Organize your data logically in a hierarchical folder structure.
- Separate active and completed work and delete any temporary files after use.
- Don't make the folder structure too nested, this can lead to long and complicated file paths.



CHECK IT OUT:

- ✓ [Biozentrum Recommendations on Proper File and Folder Names](#)
- ✓ [Basel Register of Thesauri, Ontologies & Classifications](#)
- ✓ [Digital Curation Centre](#)
- ✓ [RD Metadata Directory](#)

USE A FILE NAMING CONVENTION

Use consistent, meaningful file names.

Include the following items:

- Project (abbreviation)
- Author (whole name or initials)
- Description of the content
- Date (YYYYMMDD) / Version
- Research Team / Department

NOT all characters are allowed for folder and file naming! Do not use spaces and/or special characters in the file name.

MAX file name length: **260 CHARACTERS**

USE SUITABLE FILE FORMATS

Careful selection of **FILE FORMATS** can ensure that your files are easily accessible and interoperable and can be properly archived and still used after many years.

Aspects of a suitable Format:

- No restrictive licenses (readable by open source/code software)
- Not proprietary to one software
- No encryption or Digital Rights Management – if sharing is allowed
- Established in the research community
- Open accessible documentation

ADD DOCUMENTATION AND METADATA

The aim of the **DOCUMENTATION** is to make (permanently) stored information or documents findable again.

Structured information documenting a (data) object is called **METADATA**.

METADATA describes how the data originated and in what context it exists.

- ✓ It is recommended to define from the beginning an internal project standard how the data will be annotated.



FAIR data

 **CHECK IT OUT:**

How FAIR is your data? Take the FAIR self-assessment test!

www.andis-nectar-rds.org.au/fair-tool


The **SNSF** demands that repositories used to publish research data comply with FAIR principles. The acronym FAIR stands for:

FINDABLE, ACCESSIBLE, INTEROPERABLE, RE-USABLE

FINDABLE




Data and metadata should be easily findable for humans and computer systems. This is enabled by using machine-readable descriptive metadata.

 **DESCRIBE** provenance, usage and organization of data using standardized metadata (DataCite, RDA standards, DublinCore). Make metadata available even when data are not.

ACCESSIBLE




Data and metadata should be stored long-term so that they are easily accessible to people and machines and can be downloaded or used locally using common communication protocols.

 **OPEN** your data using standardized formats and established licences (e.g. creative commons) and state clearly what are the conditions of access to the data. Publish files in open formats alongside proprietary formats.

INTEROPERABLE




The data should be prepared so that exchange, evaluation and (re)combination with other datasets in a (semi)automated way by humans and computer systems is possible.

 **LINK** your data with persistent identifiers for datasets (e.g. DOI, HANDL, UNR) and tag all metadata with the same identifiers.

RE-USABLE



Data and metadata are well enough described that the data can be integrated in future research. Correct citation is important and the conditions and restrictions under which the data may be used should be the same for both machines and humans.

 **PUBLISH** datasets in community-supported repositories, favoring services with user-friendly interfaces.



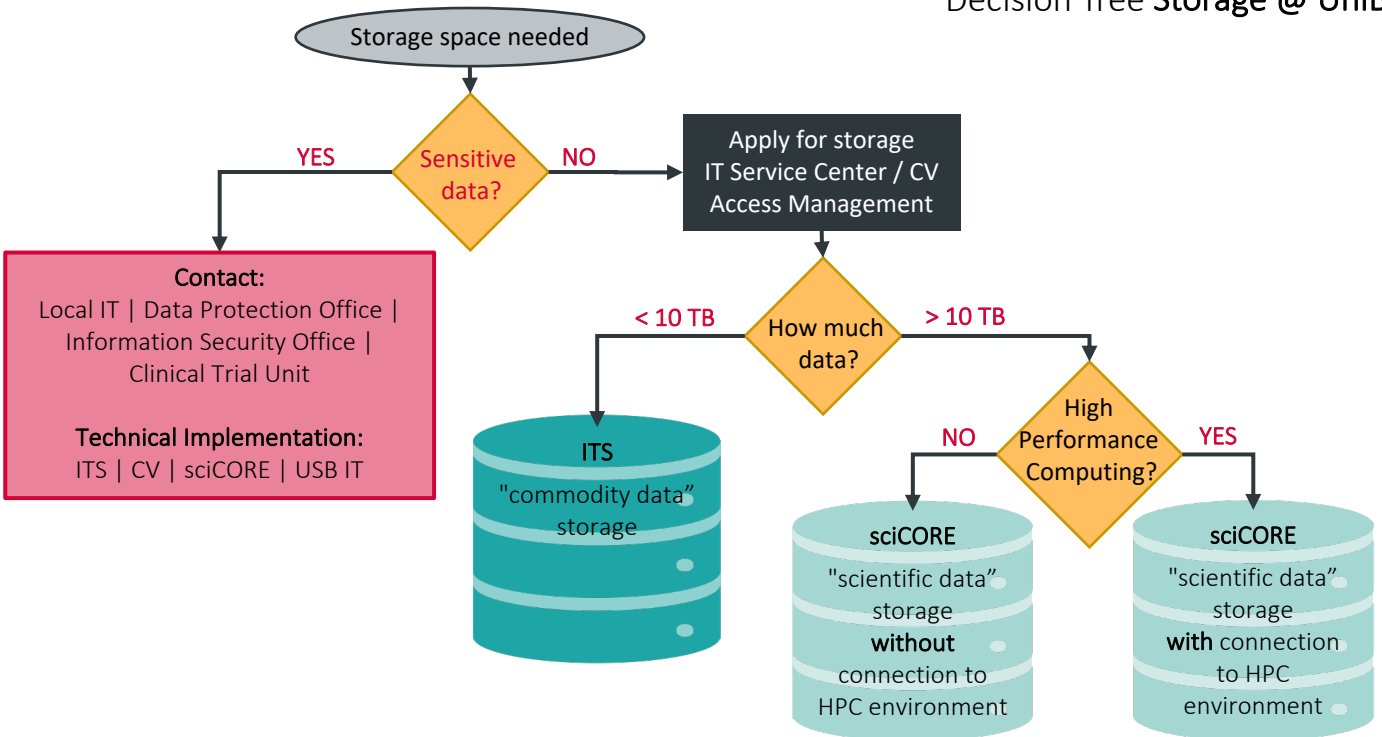
Where to store my active data?

Permanent* storage systems during active research

The decision tree helps you to find out which of the storage types offered at the University of Basel are right for you and your data.

*Does not include long-term storage, see

Decision Tree Storage @ UniBas



Document and file storage (group shares)

You can get a group share from [IT Services](#) for your daily research documents (e.g. office documents, small to mid-scale research data). For special projects with large storage requirements (tens of TBs or more) [sciCORE](#) can provide high-performance storage. Both ITS and sciCORE central storage are backed up.

Advantages of using central storage as a service:

- Professional administration
- Service availability monitored and ensured by the infrastructure provider (ITS or sciCORE)
- Infrastructure is professionally maintained
- Infrastructure is kept up to date and extended (life-cycle management)
- Established and verified disaster recovery concepts
- Data Backup Policy



Active Data Management

In the phase of **active research**, various tools and platforms are available at the University of Basel to help organize your data. This page lists key domain-specific data management services available to support your research.

NATURAL SCIENCES

Data management in natural sciences with LabKey



- [LabKey](#) is a general purpose web-based data management platform for quantitative data. It is an open-source software developed by a commercial company specialized in clinical data management. Despite being rooted in clinical data, the platform is generic enough to suit most natural (and other quantitative) sciences. LabKey provides an audit trail, a flexible role-based user management, a file manager, a configurable user interface, including a programmable wiki and is ideal when managing your research data with Excel becomes “too complicated”. Please contact [sciCORE](#) to get started.

Microscopy Image Repository with OMERO



- [OMERO](#) is a client-server software for managing, visualizing and analyzing microscopy images and associated metadata. The data repository not only supports the management of microscopy image data but also allows for their visualisation, annotation, archiving, and export of images. OMERO moreover lets the user collaborate with colleagues in the lab or anywhere in the world by creating user groups. OMERO supports over 120 different microscopy file formats. At the Biozentrum, an OMERO server is jointly operated by the department’s [Imaging Core Facility \(IMCF\)](#) and [Biozentrum Research IT](#). Please contact them for more information.

HUMANITIES & SOCIAL SCIENCES

Data management in digital humanities at DaSCH



- The [Data and Service Center for the Humanities \(DaSCH\)](#) is a national institution of the Swiss Academy of Humanities and Social Sciences (SAHSS) and the University of Basel ([Digital Humanities Lab, DHLab](#)). It offers a research data repository for the Humanities. The primary goals of the DaSCH are:
 - Preservation of research data in the humanities and their long-term data curation.
 - Ensuring permanent access to research data to make it available for further research and thus facilitating the reuse of existing research data in future research.
 - Providing services for researchers to assist them with data life cycle management.
 - Encouraging the digital interconnection and interoperability of databases created in Switzerland or abroad.
 - Collaboration and networking with other institutions on digital literacy.

Surveys with EvaSys



- [EvaSys](#), offered by the [LearnTechNet](#), is suitable for secure execution of large surveys and evaluations.



Sharing Data within a project

In the phase of active research, groups regularly come into the need for sharing their data with internal and external collaborators. This decision tree helps you decide where to place your **non-sensitive** data for safe “project-internal” sharing.

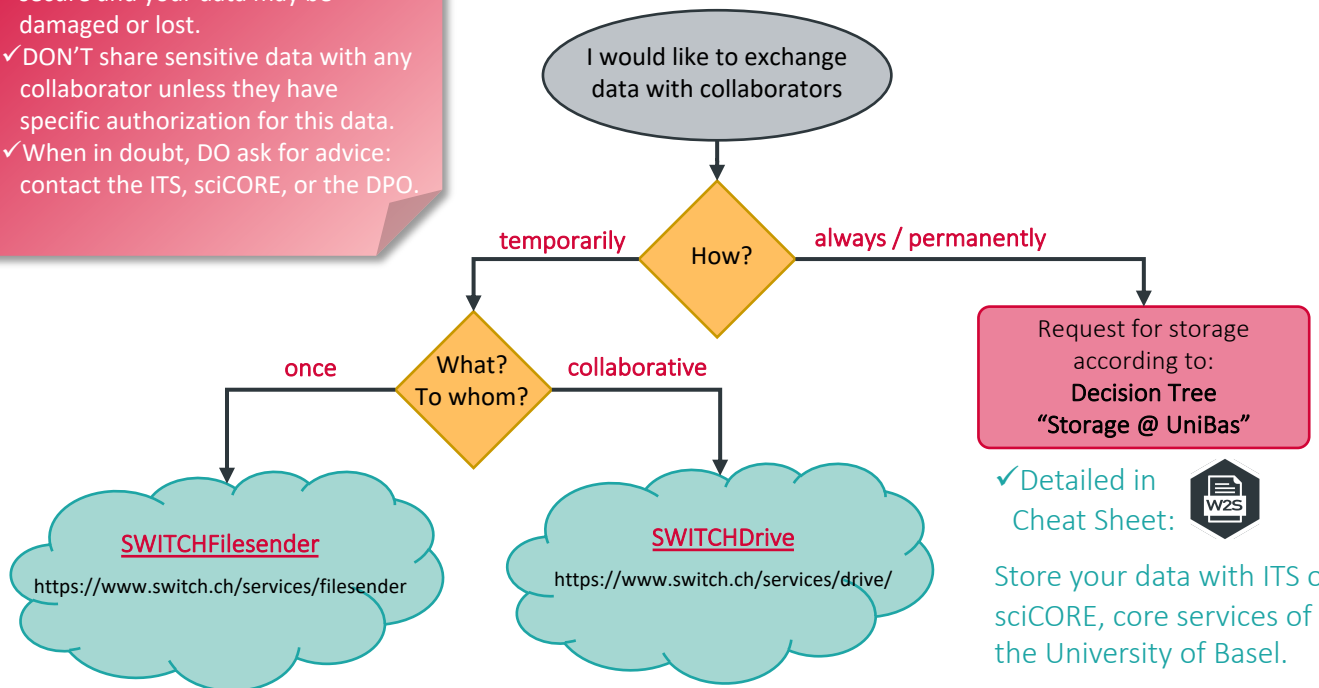
 **DO's and DONT's:**

- ✓DON'T use email, dropbox or flash drives to share your data. It is not secure and your data may be damaged or lost.
- ✓DON'T share sensitive data with any collaborator unless they have specific authorization for this data.
- ✓When in doubt, DO ask for advice: contact the ITS, sciCORE, or the DPO.

*for sensitive data check Cheat Sheets:



Decision Tree Data Exchange



✓Detailed in Cheat Sheet: 

Store your data with ITS or sciCORE, core services of the University of Basel.

Exchange large data with SWITCHfilesender

- **Filesender**: Students and University staff can use this secure web application to exchange large files with anyone:
 - Maximum number of files to upload per shipment: 1 (Tip: ZIP archive)
 - Maximum file size per upload: 50 GB
 - Maximum number of e-mail recipients: 100
 - Maximum storage time for files: 20 days

Secure cloud service for Swiss universities with SWITCHdrive

- **SWITCHdrive** (cloud storage) enables collaborative work, as users can share their files. It provides 50GB of storage space and requires Swiss educational affiliation (SwitchAAI edu-ID). Users can access their data from their desktops, smartphones and tablets.

Hold online meetings and webinars with SWITCHinteract



Publishing Data



In parallel to a publication in a peer-reviewed journal, **make the (raw) data available to the research community**. This allows other researchers to **reproduce** your findings and enables **reusability** in subsequent research.

When you finish your study, **make your data accessible**. Here you find useful information to help you **publish** and **disseminate** your data.

 Before sharing your data, make sure you are allowed to! Beware of restrictions in sharing sensitive data!



REPOSITORIES

Domain-specific Repositories

- [Re3data](#): a catalog of repositories
- [Deposition Databases for Biomolecular Data](#) by Elixir
- Subject related repositories overview by:
 - the [Open Access Directory](#)
 - [PLOS](#)

Multidisciplinary repositories

- [ZENODO](#) is a research data repository created, hosted and operated by CERN and OpenAIRE
- [DRYAD](#) is a research data repository, originally funded by the National Science Foundation, governed by a nonprofit membership organization
- Recommendations of the Swiss Academy of Humanities and Social Sciences:
 - Humanities - the [SAGW](#) repository
 - - the [DaSCH](#) platform
 - Social sciences - the [FORS](#) repository

! *Data published in Zenodo are assumed to be open, the service is thus unsuitable for sensitive data, such as personal or copyrighted data.*

! *SNSF grants require the use of not-for-profit storage solutions for data. Dryad and Zenodo are compliant with this request. Figshare, another common repository, is operated by Macmillan Publishers and therefore not allowed.*

FINDING DATA

- (Multi)disciplinary archives can be found with the help of [Re3data](#) – Registry of Research Data Repositories.
- Some data sets are findable via Google, tools or customized search engines for data, such as:
 - [b2find.eudat.eu](#)
 - [datacite.org](#)

CITING DATA AND IDENTIFIERS

Persistent identifiers allow scientists and repositories to uniquely identify persons or cite datasets.

- [DOI](#): **Digital Object Identifier**
A unique alphanumeric string assigned by the International DOI Foundation to identify content and provide a persistent link to its web location.
- [ORCID](#): **Open Researcher and Contributor Identifier**
It provides an identifier for authors as they engage in research, scholarship, and innovation activities.
- [ARK ID](#): **Archival Resource Key Identifier**
A URL designed to support long-term access information objects.

DATA INTEROPERABILITY

- [BIOSCHEMAS](#) is a community effort to improve data interoperability in life sciences. It encourages people in life science to use [schema.org](#) markup, so that their websites and services contain consistently structured information.



Sensitive Data Management

Laws and regulations



Which data is **sensitive**? What can you do (and not do) with it? Does your research project involve sensitive data? Do you know **how to handle** it and **where to store** it?

Any researcher working with sensitive data needs to be aware of the relevant **laws** and **regulations** and is subject to the (swiss) data protection law.

DEFINITIONS

- **PERSONAL DATA:** Any information relating to an *identified or identifiable natural person* ('data subject') e.g. **Name | ID number | Location data | Online identifier | Factors of physical, physiological, genetic, mental, economic, cultural or social identity.**
- **SENSITIVE DATA:** Broader scope which includes but is not restricted to personal data e.g. **Political, religious or philosophical beliefs | Memberships | Company secrets | Contracts | Private communications | Processing of genetic / biometric data.**

KEY LAWS

The **CANTONAL LAW ON INFORMATION AND DATA PROTECTION (IDG-BS)**, the **SWISS FEDERAL ACT ON DATA PROTECTION (FADP)** and the **EU GENERAL DATA PROTECTION REGULATION (GDPR)** protect privacy and regulate the handling of personal data.

THE HUMAN RESEARCH ACT (HRA) protects the dignity, privacy and health of human beings involved in research.



IN SWITZERLAND: Which law applies when you process personal data?

- **Research at the University of Basel:** In general, the Cantonal Law on Information and Data Protection **IDG-BS** applies to public bodies of the canton Basel-Stadt and its municipalities (e.g. University of Basel).
- **Private Research/Research at the ETH:** In general, the **SWISS FEDERAL ACT ON DATA PROTECTION (FADP)** and **the Ordinance To The FADP** apply to private individuals and federal bodies (e.g. ETH).
- **Cross-cantonal research:** In case of processing personal data usually, the data protection law of the canton in which the processing appears (e.g. collect, store, revision, disclose) the data is applicable. That means that more than one cantonal data protection law can apply.
- **Research with an international dimension:**
 - ! Often the data protection law at the foreign research location is also applicable.
 - ! If you process personal data of EU individuals or monitor the online behavior of users based in the EU, **GDPR** is applicable.
 - ! **Best practice: ask for legal advice.**

CONTACT

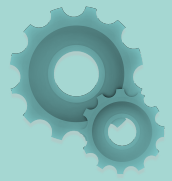
In case of doubt, just ASK:

- The **Data Protection Officer (DPO)** of the University of Basel: Danielle Kaufmann | danielle.kaufmann@unibas.ch
- The **Information Security Officer (CISO)** of the University of Basel : Ernestine Schikore | ernestine.schikore@unibas.ch
- The University of Basel **Research Data Management Network:** researchdata@unibas.ch



Sensitive Data Management

Data processing



Processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

Main principles of data protection*

LAWFULNESS: either on the basis of a **legal regulation** and/or with the **consent of the data subject**, or due to an overriding **legitimate interest** of the processing person.

EARMARKING: Data processing must always be carried out for a **specific purpose**.

PROPORTIONALITY: Data processing must be **necessary for the intended purpose** and reasonable in relation to the infringement of privacy.

CORRECTNESS (INTEGRITY): Whoever processes personal data must ensure that the data is **correct**.

PERCEPTABILITY OF DATA PROCESSING: It must be **clear** to the data subject that personal data relating to them are collected and processed.

TRANSPARENCY: Data subjects must be adequately **informed** about data processing so that they can understand what is done with their data and for what purpose. The data subjects also have the right at any time and without giving reasons to request information about their data.

DATA SECURITY: The data processing must comply with technical and organizational security requirements.

*Fact Sheet Datenschutz –
Wichtigste Definitionen & Beispiele
<https://tinyurl.com/FS-Datenschutz-DB>



Informed Consent*

Whether consent is mandatory or supplementary to a legal basis or sufficient as a justification depends on the **applicable data protection law**. For concrete data processing an informed consent is always required.

The **declaration of consent** must:

- contain type and scope of the data and data processing
- contain a reference to the intended use
- contain possible disclosure of the data to 3rd parties
- list the rights of the data subject to cancellation, access and opposition
- be voluntary; the data subject must not suffer disadvantages as a result of refusing consent.

Consents can fulfil several functions, including the provision of information (**principle of transparency**) and can specify a general legal basis.

*Fact Sheet Einwilligung
<https://tinyurl.com/FS-Einwilligung>



Sensitive data in the University of Basel

- If you work with sensitive data, always contact the **IT responsible** of your department.
- Get useful tips on the **ITS Data Security** website: its.unibas.ch/sicherheit
- For complex issues: the **DPO** Danielle Kaufmann | danielle.kaufmann@unibas.ch and the **CISO** Ernestine Schikore | ernestine.schikore@unibas.ch
- **sciCORE** offers a **secure** space for storing sensitive data: scicore.unibas.ch
- **BioMedIT**: a coordinated nationwide network of secure infrastructures: scicore.unibas.ch/projects/biomedit/
- **Biozentrum Research IT** | researchit-biozentrum@unibas.ch



Sensitive Data Management Data masking



Often you will be required by law to de-identify your data in order to be able to use it. This guide explains some of the basic terms regarding data **de- and re-identification** and provides practical examples.



Useful links:

- ✓ arx.deidentifier.org/
- ✓ qosient.com/argus/anonymization.shtml
- ✓ cran.r-project.org/web/packages/sdcMicro/index.html

DEFINITIONS

- **ANONYMISATION** means that the reference to a person is irreversibly (= definitively) removed in such a way that it is no longer possible to draw conclusions about persons without disproportionate effort. Anonymised data are no longer regarded as personal data.
- **PSEUDONYMISATION** refers to the removal of personal references, whereby a specific key (i.e. a table with the translation pseudonym to person) is retained for the re-personalisation of the information. If data is pseudonymised, the conditions under which a person may be identified and how the key is stored must be regulated (key management). Unlike anonymised data, pseudonymised data remain personal data.
- **ENCRYPTION** is the process by which a plain text is converted into a non-interpretable character string (ciphertext) by means of an encryption procedure (cryptosystem). One or more keys (=codes) are used as crucial parameters for encryption. Secure encryption makes it practically impossible to retrieve the plain text without the appropriate key. However, secure encryption must be done correctly. When in doubt, contact your local IT.

EXAMPLES

Collected data	ID	Name	Date of birth	Disease	Treatment
	DS001	Alan Smith	01.06.2001	Appendicitis	Operation
Anonymised data	ID	Name	Date of birth	Disease	Treatment
	DS001			Appendicitis	Operation
Pseudonymised data Research Table	ID			Disease	Treatment
	DS001			Appendicitis	Operation
Mapping Table (Key)	ID	Name	Date of birth		
	DS001	Alan Smith	01.06.2001		
Encrypted data*	ZTUVqwEAbUf +M0tKeQXzQ VjnI3VMMvJT dpgB9eIMbzs	xu05js8QCrn90lj1u afISzUybf8uMEvrD T+KSXvJY24	IYtVtyocsRqTI6QAN xqfDcjxxvRHSWEGJ QY54ddSvZE	TOP2AMUh/Fgs2Xh oCY2Zv30qaKPwZ/ 4A7t1FpLcRBng	P7Hz796XwSkPxEQ vGv2PGcR8XUmG/I C0kSkv6n/bnlc
	FzNimG2xuPF M9mnsFFoEp YEUXR+z7mzn yyK7Ka9SEbw	aq9pIVYmGCPF+Hp I+dlamkEpAV/n8alc Nz/Cb1d31L4	MMC/q92DGxcz9v 6fKycELg+IK6N/575 SSU1SMR1uKWc	/o9WGGyX+jQl2LYv GTweb1At84gJVVD EgjdYFsbVkom	GUSE+ZAuKrDN+X9 zduJmhNiCrLGUL+s Ex//EzLHtXq8

* Online-toolz encryption tool used for this example only! *Online encryption tools are NOT suitable to use for sensitive data!*



Moving on

What to do before leaving the university?

In case you are leaving the University of Basel you have to think about a few questions related to your data sets. In order to get your data **transferred** and/or **stored** properly before leaving we recommend to start **planning** a few months ahead of your departure.

Here is what you have to do:

STEP 1 CLARIFY AND CHECK CONDITIONS THAT APPLY TO YOUR DATA SET

Find out or ask yourself some of the following questions:

- Is your data part of a publication and do you want or have to publish it?
- Does your data set fall under the integrity rule of the university and have to be stored for min 5 years?
- Do you have data on central storage infrastructure* that should be cleared? * e.g. Tucan of sciCore or HNAS Server

STEP 2 GET SUPPORT AND DISCUSSION PARTNERS

Get in touch with your local IT or storage provider where you have your data stored. Discuss where the data should go and options to achieve this. Depending on:

- legal restrictions that apply
- the storage conditions and access needs
- your financial capabilities

choose from the following paths:

DATA STAYS at the University

DATA LEAVES with the researcher

STEP 3 DECIDE WHAT PART GOES WHERE

PRESERVATION POSSIBILITIES

on tape on bulk storage operated by

- IT Services
- University Library

PUBLICATION POSSIBILITIES

- Subject specific repositories
- General repositories

STEP 3 DEVELOP PLAN FOR TRANSFER WITH

- your local IT specialist
- department manager
- legal service



Useful time estimates for planning transfers of large data volumes

Task	Month 1	Month 2	Month 3	Month 4	Month 5
Packing of data	█				
Transfer of data		█	█	█	
Purchase of shuttle	█				
Delivery of shuttle		█	█	█	
Shuttle set up			█		
Data synchronisation			█	█	█
Benchmark/tune transfer	█				
Shuttle			█	█	
Transport					█
Setup at destination					█

