


Forschungsdatenmanagement - was kann ich tun?

Educational Material

Author(s):

Petrus, Ana 

Publication date:

2018-03-20

Permanent link:

<https://doi.org/10.3929/ethz-b-000297435>

Rights / license:

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International



Forschungsdatenmanagement – Was kann ich tun?

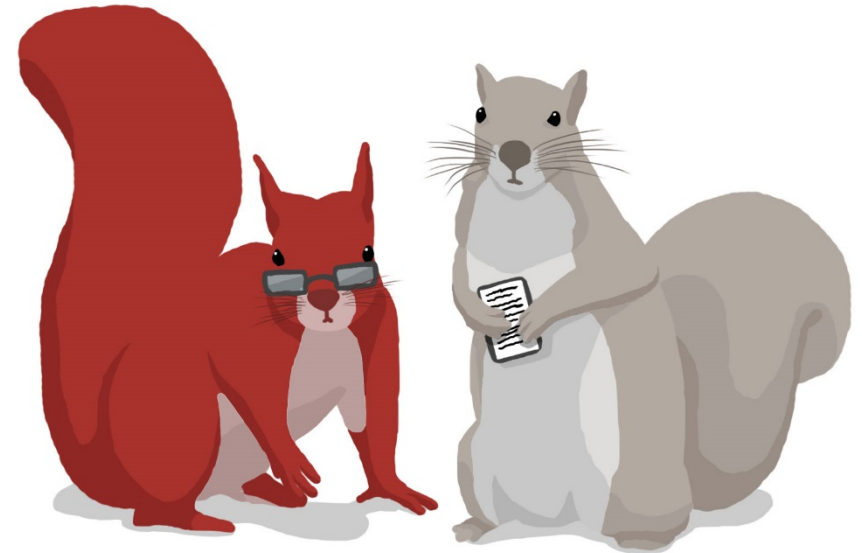
Dr. Ana Sesartic – Digitaler Datenerhalt

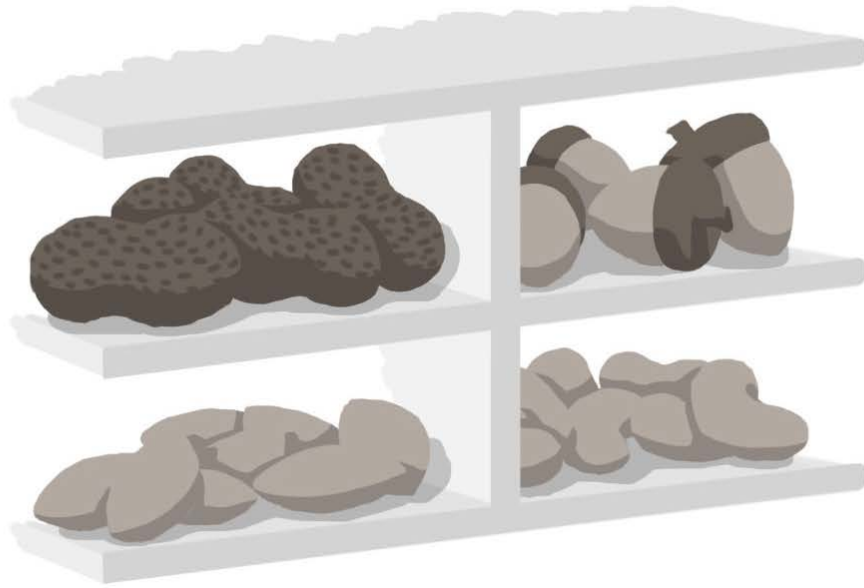
Heutiges Programm

1. Was ist Datenmanagement?
2. Vorschriften der ETH Zürich
3. Was ist ein Datenmanagementplan?
4. Vorbereitung auf die Langzeitarchivierung
5. Tipps und Tricks zur Organisation
6. Blick in den Werkzeugkasten
7. Dienstleistungen der ETH Zürich

Erzählen Sie kurz über

- Ihren (wissenschaftlichen) **Hintergrund**,
- bisherige **Erfahrungen** mit Datenmanagement und
- Ihre **Motivation** zur Kursteilnahme.





Was ist Datenmanagement?

Datenmanagement bedeutet

- die **Organisation**,
 - **Strukturierung**,
 - **Speicherung** und
 - **Pflege Ihrer Daten***
- über die **Laufzeit** des Projektes und
 - nach dessen **Abschluss**.

*Daten sind alle Dokumente, die Informationen auf eine interpretierbare Weise widerspiegeln. Dazu gehören Text, Bild, Video, Audio bis hin zu Software.

Weshalb betrifft es uns?



Gut geführtes Datenmanagement

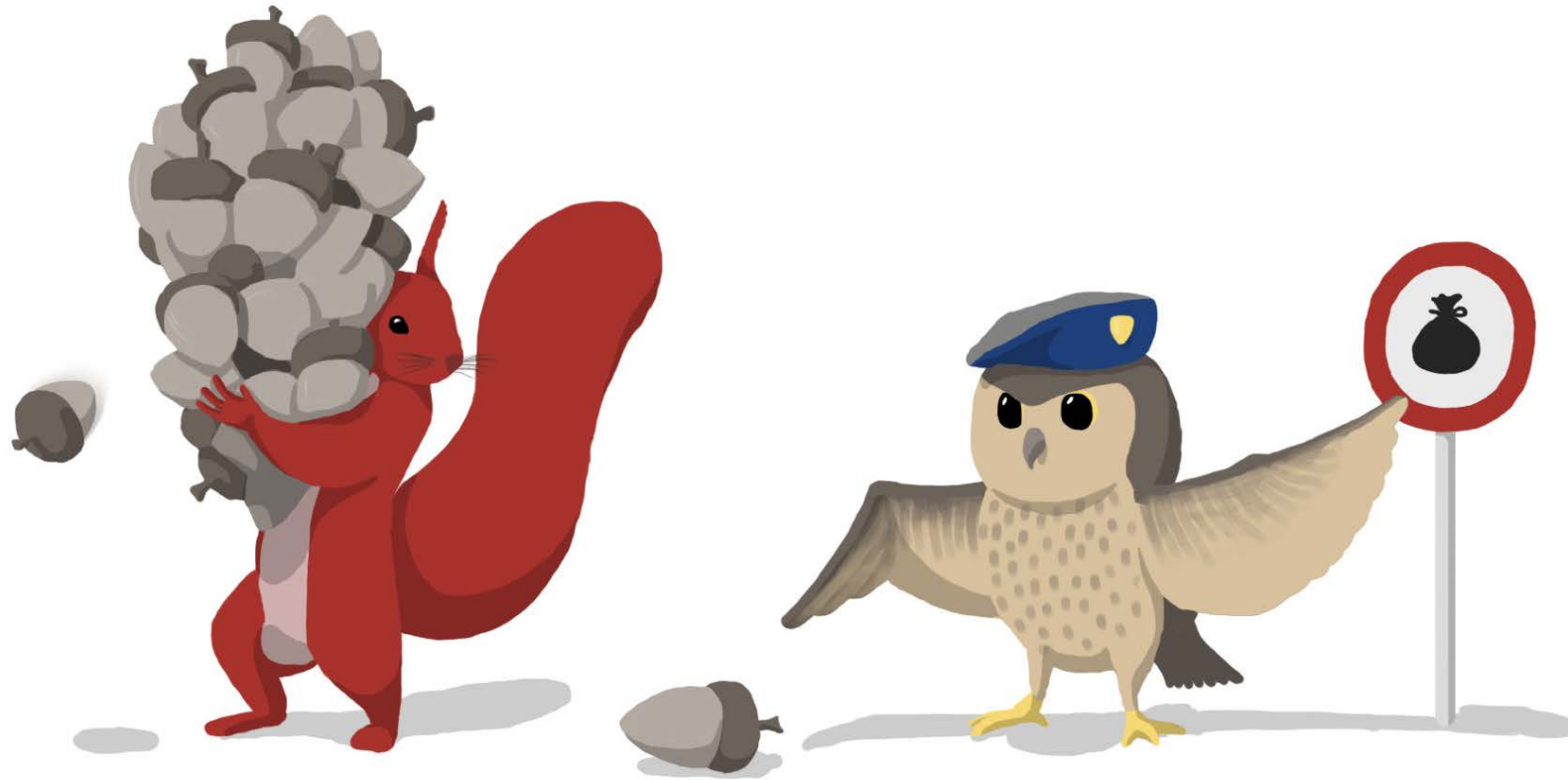
- Erhöht die **Effizienz** Ihrer Forschung
- Erleichtert das **Teilen** und **Veröffentlichen** Ihrer Daten
- Macht Ihre Arbeit auch in Zukunft **nachvollziehbar**
- Ist **gute wissenschaftliche Praxis** und erhöht die Transparenz und Validität
- Gehört zu den Vorgaben von **Arbeit- und Geldgebern***

*[SNF](#) und [Horizon 2020](#) fordern Datenmanagementpläne

So lieber nicht...



- Daten erstellen, **ohne** sich **über** ihre **Veröffentlichung Gedanken zu machen.**
- **Keine** umfassende **Dokumentation** der Forschungsdaten führen.
- Nur technische Metadaten können später extrahiert werden, aber **wenig bis gar keine Dokumentation von Inhalt oder Kontext** kann sinnvoll hinzugefügt werden.



Vorschriften der ETH Zürich zu geistigem Eigentum, Privatsphäre und Zugriffsrechte

MANAGE YOUR DATA



Protect your brainwork. IT Security Initiative

Entfalten Sie das volle Potential Ihrer Daten und schaffen Sie die perfekte Voraussetzung für effizientes, produktives und sicheres Arbeiten. Wie? Mit solidem Datenmanagement. Finden Sie hier die wichtigsten Tipps, Hinweise und Links zu erfolgreichem Datenmanagement.

itsecurity.ethz.ch/de/#/manage_your_data

Richtlinien für Integrität in der Forschung



«An der ETH Zürich beruht Forschung auf **intellektueller Redlichkeit**. Die Forschenden [...] stehen für **wissenschaftliche Integrität** und **Wahrhaftigkeit** in der Forschung und bei Begutachtungen.»

doi.org/10.3929/ethz-b-000179298

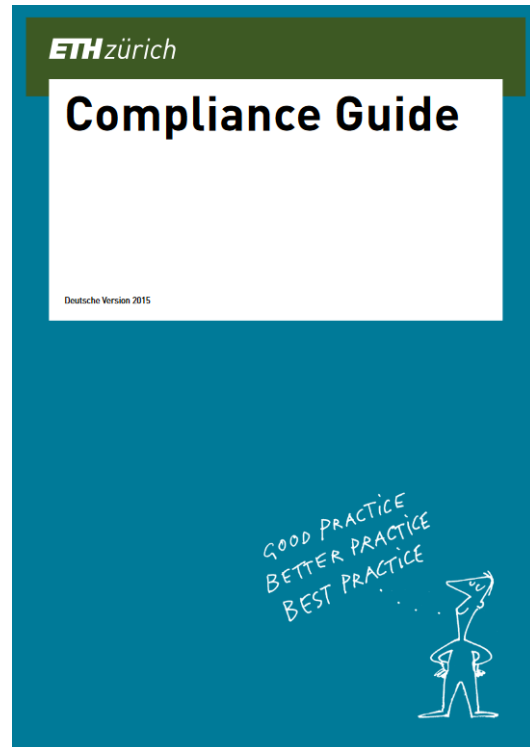
Artikel 11.

Erhebung, Dokumentation und Aufbewahrung von Primärdaten

«Sämtliche **Verfahrensschritte** im Umgang mit den Primärdaten (statistische Analysen, Umformungen etc.) sind **in einer der jeweiligen Disziplin angepassten Form** (z.B. Laborjournal, andere Datenträger) derart zu **dokumentieren**, dass die aus den Primärdaten gewonnenen **Ergebnisse vollständig reproduziert werden können.**»

«Die **Projektleitung ist für das Management der Daten** (Aufbewahrung, Datenzugang, Einhaltung des Datenschutzes, etc.) **verantwortlich**. Sie sorgt insbesondere dafür, dass **Daten und Materialien** nach Abschluss des Projektes **während der für das Fachgebiet massgebenden Frist aufbewahrt bleiben** und gegebenenfalls innerhalb der gesetzlich vorgegeben Frist ordnungsgemäss vernichtet werden.»

Compliance Guide



«[...] **alle [ETH] Angehörigen** [sollen sich] an die gesetzlichen Vorschriften und internen Weisungen halten und diese in ihrem Bereich umsetzen.»

«Der vorliegende Compliance Guide bietet dazu eine **Orientierungshilfe**. [...] Damit die Umsetzung leichter fällt, sind zu jedem Punkt weitere Informationskanäle und Auskunftspersonen angegeben, die konsultiert werden können.»

rechtssammlung.sp.ethz.ch/Dokumente/133.pdf

Wissen Sie wo Ihre Daten sind und wer Zugriff hat?



There is NO CLOUD, just other people's computers

["There is no cloud just other people's computers"](#)
by Markus Meier
[CC BY-SA 4.0](#)

Cloud Computing @ ETH Zürich

Regeln und Vorschriften

«Die Auslagerung von sensitiven Daten der ETH Zürich (z.B. Forschungsdaten, die einer vertraglichen Geheimhaltung mit Dritten unterliegen, wichtige Geschäftsdaten der ETH Zürich wie etwa Finanzdaten, personenbezogene Mitarbeiter- oder Studierendendaten, Gutachten) ist nicht zulässig. Die **ETH Zürich muss jederzeit** den **Zugriff** und die Kontrolle über diese Daten **haben.**»

«Die Nutzung von **Cloud** oder Social Media Services (z.B. Facebook, Google, Dropbox) in der Forschung für den Austausch mit Forschenden anderer Hochschulen, in der Lehre für den Austausch mit den Studierenden (Vorlesungsfolder, etc.) ist unproblematisch, solange **keine sensitiven Daten** der ETH Zürich betroffen sind und keine Rechte Dritter, namentlich Persönlichkeits- oder Urheberrechte, verletzt werden.»

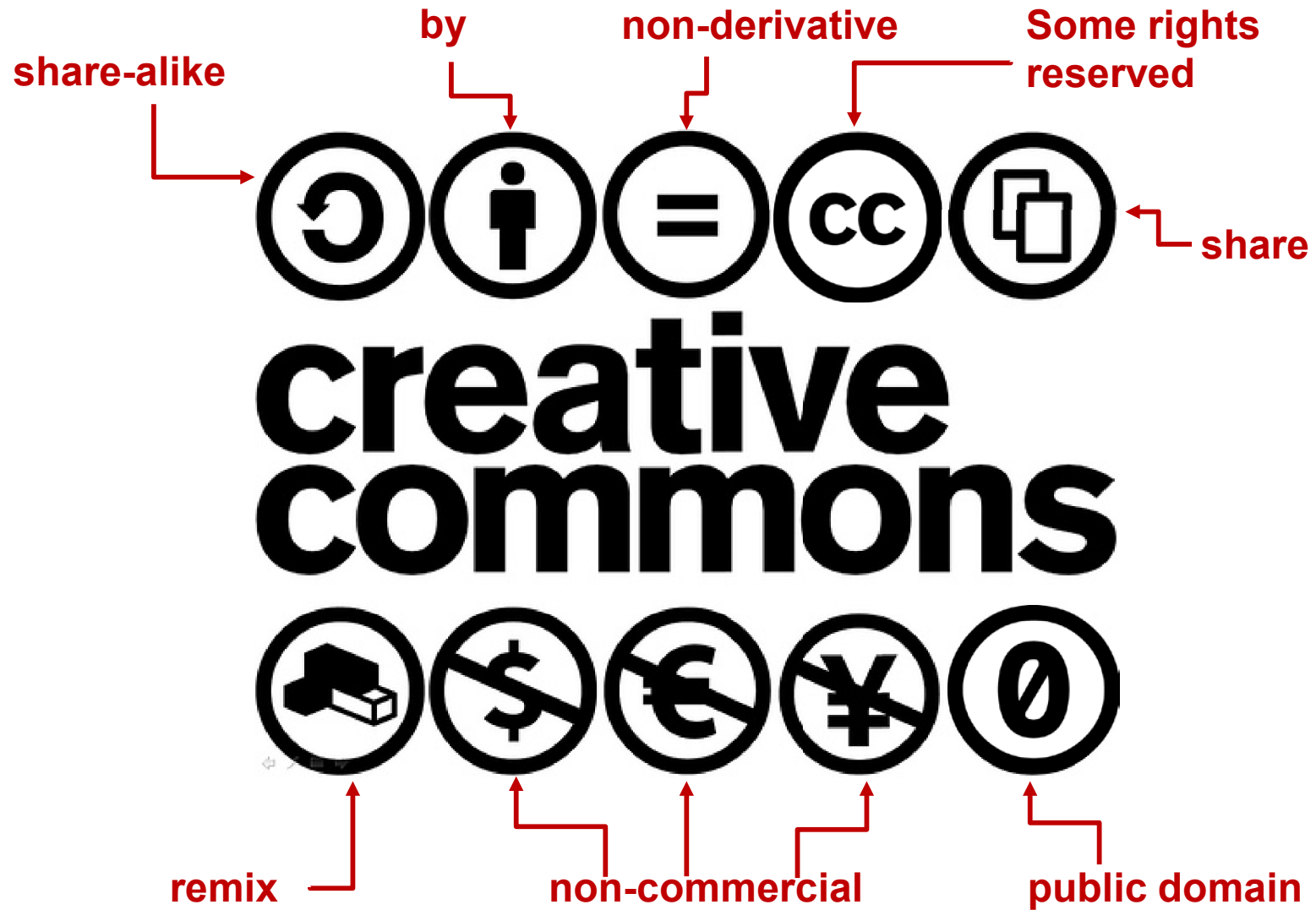
www.ethz.ch/content/dam/ethz/associates/services/organisation/departments/informatikdienste/files/Projekte/leaflet_example_cloud_DE.pdf

Privatsphäre

- Personenbezogene Daten sind nach **schweizerischem Datenschutzrecht** aufzubewahren
- Eine entsprechende **Anonymisierung** kann erforderlich sein
- Das **Löschen** einzelner Datensätze muss jederzeit möglich sein
- Die Testpersonen müssen eine **Einverständniserklärung** unterzeichnen

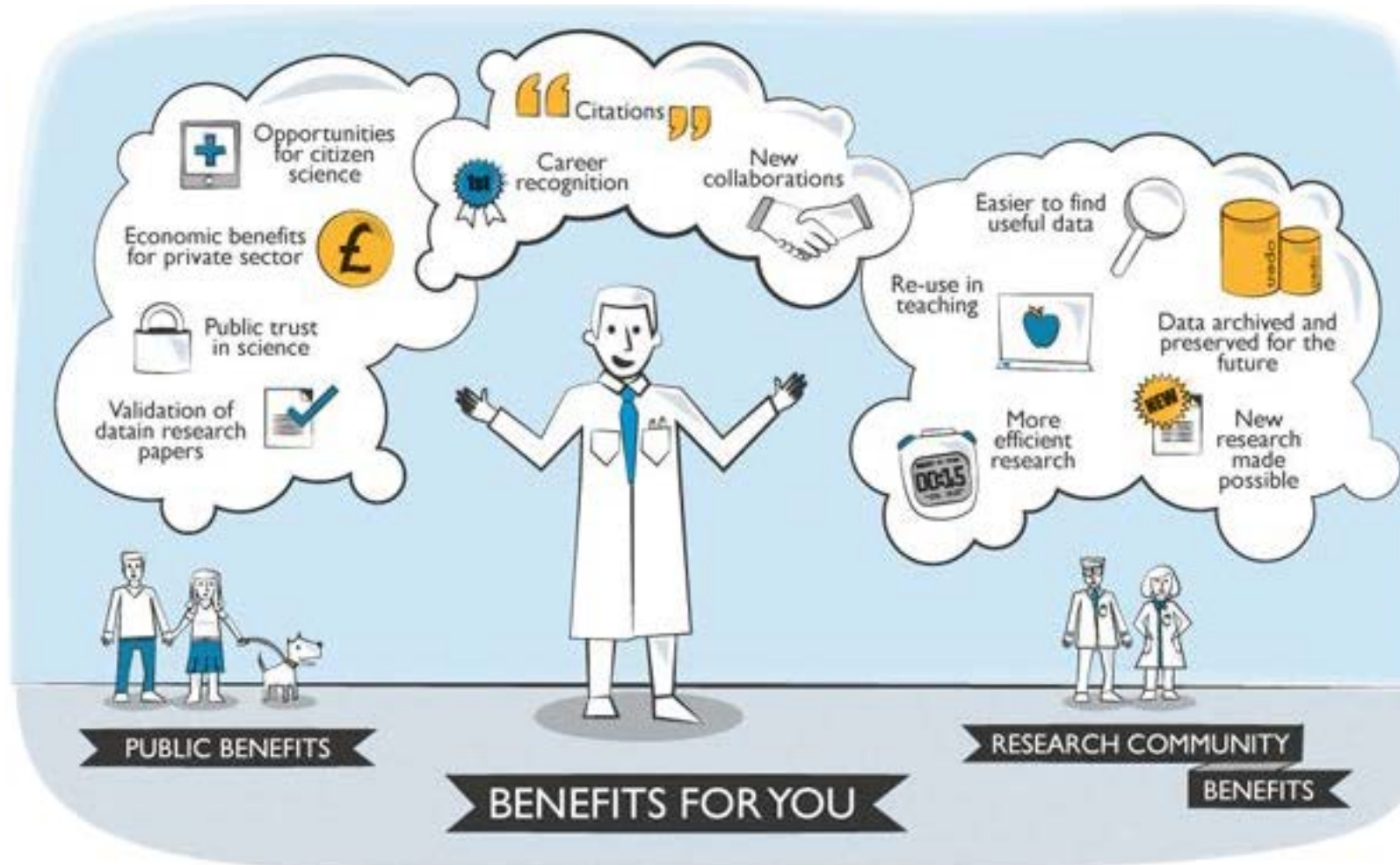
Geistiges Eigentum

- **Respektieren** Sie die Rechte anderer (Dritte, Mitarbeitende)
- Im Zweifelsfall auch bei einer CC-Lizenz **um Genehmigung fragen**
- Die meisten **immateriellen Rechte an Werken gehören der ETH Zürich**.
Im Zweifelsfalle an ETH transfer wenden (www.transfer.ethz.ch)
- Achten Sie darauf, **ausreichende Rechte zu behalten**
 - z.B. für Open Access Publishing («grüner Weg», www.library.ethz.ch/Open-Access)
 - z.B. in Bezug auf Patentanmeldungen: ETH transfer (www.transfer.ethz.ch)



["Creative Commons"](#) (4.9.2018)
by Michael Porter
[CC BY-NC-ND 2.0](#)

Vorteile des Teilens von Daten



(Wieder-) Verwendung von Daten kritisch überdenken

- Daten, Metadaten und Kontext werden benötigt, um einen Datensatz richtig zu verstehen.
- Das Datenmanagement beinhaltet nicht nur Ihre eigenen Daten, sondern auch eine kritische Sicht auf fremde, von Ihnen verwendeten Daten:
 - Verstehen Sie, wie die Daten gewonnen wurden?
 - Haben Sie genügend Informationen, um ihre Zuverlässigkeit zu bewerten?
 - Können Sie die Daten verwenden, ohne mit ihren Urhebern Rücksprache halten zu müssen?
 - Werden Sie in ein paar Monaten noch wissen, welche Daten Sie von anderen Forschern wiederverwendet haben?
 - Wissen Sie, wie Sie die von Ihnen verwendeten Daten zitieren sollen? (siehe www.datacite.org/cite-your-data.html)



Was ist ein Datenmanagementplan?

Der Datenmanagementplan ist

ein kurzer Plan, der zum **Projektstart** angelegt und **während des Projektverlaufs aktualisiert wird**. Er beantwortet folgende Fragen:

- **Welche Daten** werden gesammelt oder kreiert?
- Wie werden die Daten **dokumentiert und beschrieben**?
- **Wo** werden die Daten **gespeichert**?
- **Wer** ist für die Sicherung der Daten **verantwortlich**?
- Welche Daten sollen **geteilt und/oder langzeitarchiviert** werden?
- **Wie und mit wem** werden die Daten geteilt?

Weshalb ein DMP?

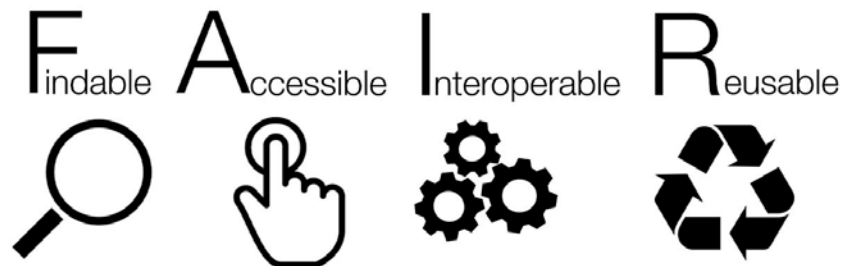
Sie werden zunehmend für Projektanträge verlangt (SNF, Horizon 2020) und helfen Forschenden bei Organisation und Pflege ihrer Daten. Ein DMP hilft Ihnen

- **bewusste Entscheidungen zu treffen**, um Probleme zu vermeiden,
- Prozeduren für die Erstellung von **konsistenten Daten frühzeitig zu entwickeln**,
- sicherzustellen, dass Daten **korrekt, vollständig, zuverlässig und sicher** sind,
- (unerwünschte) **Duplizierung, Datenverlust und Sicherheitsverletzungen** zu vermeiden,
- **Zeit und Mühe** zu sparen.

Kurzum: Ein DMP erleichtert Ihr Leben!

Grundsätze des SNF zu Open Research Data

- Forschungsdaten sollen für Forschende wie auch Laien **frei zugänglich sein**.
- Der DMP hilft Forschenden bewusst zu entscheiden, **welche Daten offengelegt** und **welche langzeitarchiviert** werden sollen.
- Daten müssen den **FAIR-Prinzipien** folgen.



[FAIR image](#) (4.9.2018) by Sangya Pundir / [CC BY-SA 4.0](#)

SNF-Begutachtung des DMP

- Die **Plausibilität** und **Übereinstimmung** mit der SNF Policy zu Open Research Data werden geprüft.
- Der DMP wird **NICHT** an externe Gutachter geschickt.
- Akzeptierten Anträgen können **Auflagen zur Verbesserung des DMP** gemacht werden.

DMP Guidelines, FAQ etc. www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx

Kontakt-Adresse für Fragen, Vorschläge und Kommentare: ord@snf.ch

Der DMP bei SNF

- **muss** online über **mySNF** kreiert,
- im Projektverlauf **aktualisiert** und
- zum Schluss **auf p3.snf.ch veröffentlicht** werden.

www.mysnf.ch

Inhalte des DMP gemäss SNF:

www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_de.pdf

Hilfe und Beispiele der ETH Zürich (nur auf Englisch):

documentation.library.ethz.ch/display/DD/Guidance+for+ETH+researchers+on+filling+out+SNSF+Data+Management+Plans

The screenshot shows the mySNF login interface. At the top left, there are logos for SNF (Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung) and mySNF. On the right, there are navigation links for 'D', 'F', and 'E'. A prominent red callout box with white text is overlaid on the right side, reading: 'Wir bieten Workshops und Beratungen an. Kommen Sie auf uns zu!'. The main content area is titled 'Willkommen bei mySNF' and contains introductory text about the platform's purpose and a login form. The login form includes fields for 'Benutzername' and 'Passwort', an 'Anmelden' button, and links for 'Benutzer/in neu registrieren' and 'Zugangsdaten vergessen'. Below the form, there are 'Weiterführende Links' for user registration and the mySNF usage policy.

Informationen bei anderen Geldgebern

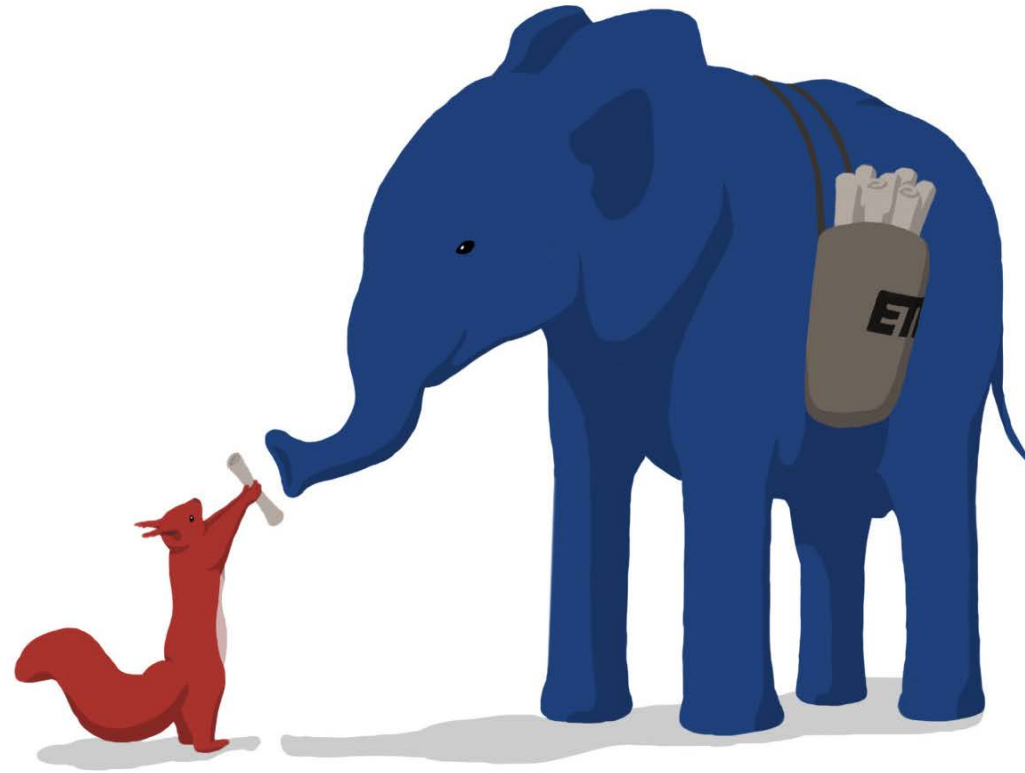
Checkliste der ETH-Bibliothek

- Informiert über Datenmanagement
- Hilft bei der Erstellung des DMP

documentation.library.ethz.ch/display/DD/Data+Management+Checklist

DMPonline vom UK Digital Curation Centre

- Hilfreiches Werkzeug für die Erstellung eines DMP für **Horizon 2020**
dmponline.dcc.ac.uk
- Beispielsammlung
www.dcc.ac.uk/resources/data-management-plans/guidance-examples



Vorbereitung auf die Langzeitarchivierung

Vom Datenmanagement zur Langzeitarchivierung

- Gutes Datenmanagement vereinfacht den **Übergang** zur Langzeitarchivierung.
- Sollen Daten über ein Jahrzehnt erhalten bleiben, sollten sie archiviert werden.
- Weitere Massnahmen zur Langzeitarchivierung:
 - **Daten möglichst selbsterklärend sichern**, einschliesslich Dokumentationen, Software, und z.B. Referenzoutputs von Modellalgorithmen
 - sorgfältige Auswahl von **Dateiformaten**

Dateiformate sollten

- **offenen** Standards entsprechen (nicht proprietär),
- gut **dokumentiert**,
- **weit verbreitet** und durch viele Programme **unterstützt**,
- **unkomprimiert** (oder min. verlustfrei komprimiert) und **unverschlüsselt** sein.

Im Zweifelsfall **Original beibehalten und eine Kopie** in einem offenen Format erstellen.

Beispiele

- **Bilder:** unkomprimiertes TIFF; JPEG2000
- **Text:** ASCII, inklusive XML etc.*
- **Formatierter Text:** PDF/A1-b, (PDF)
- **Daten aus Tabellen:** CSV
- **Tabellen:** (CSV), (ODF, OOXML)

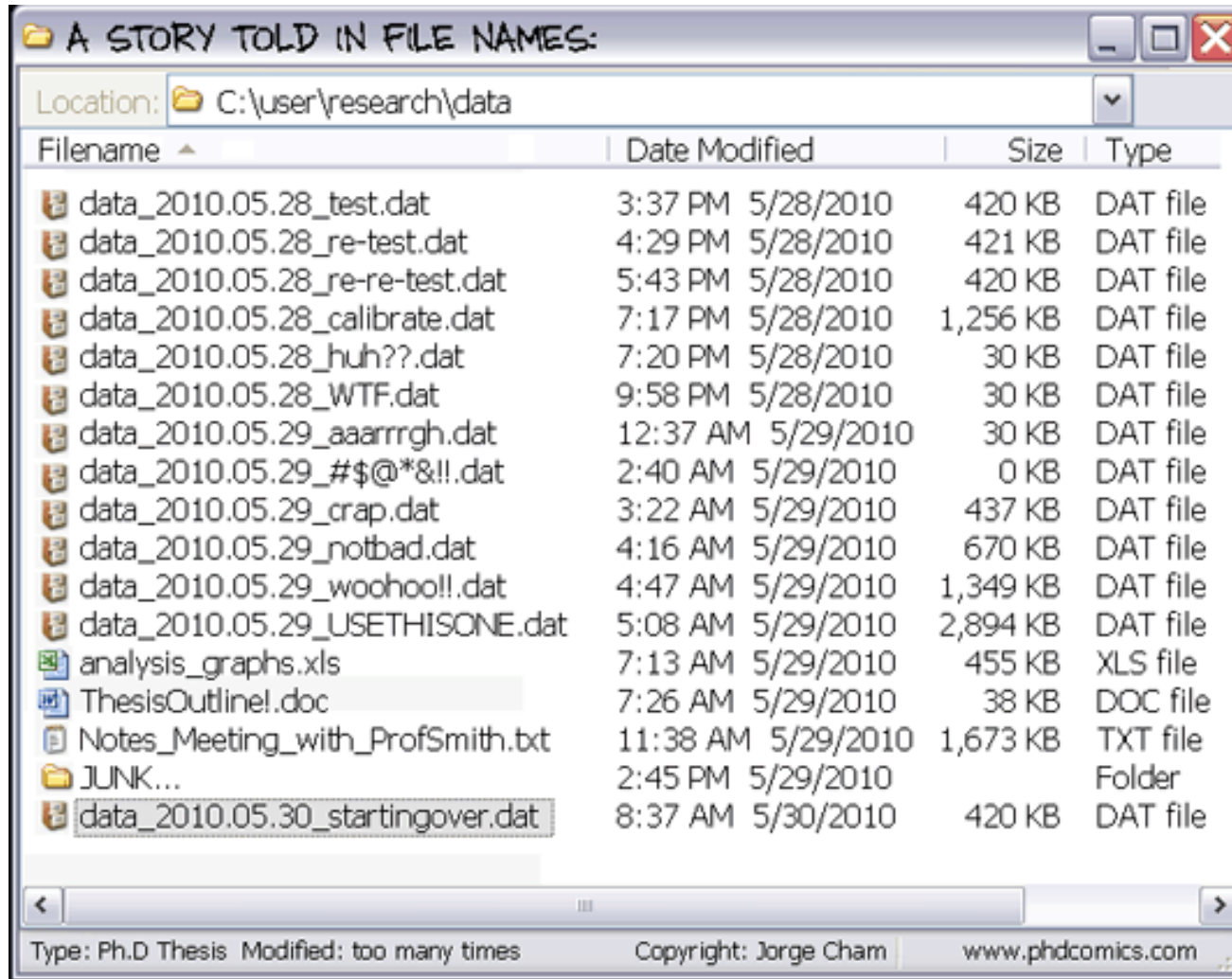
*Informationen über **Kodierung** und **Abhängigkeiten** wie Stylesheets und TeX-Bibliotheken nicht vergessen

Aber das bedeutet nicht,

- dass Sie Daten nicht in anderen Formaten abspeichern dürfen. Sie müssen sich **dessen nur bewusst sein** und sich **auf die Zukunft vorbereiten**.
- Überlegen Sie sich **alternative Formate** (ja, redundant!) zu den proprietären Formaten zu benutzen.



Tipps und Tricks zur Organisation



A STORY TOLD IN FILE NAMES:

Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

Type: Ph.D Thesis Modified: too many times Copyright: Jorge Cham www.phdcomics.com

Kommt Ihnen das bekannt vor?

"Piled Higher and Deeper" by
Jorge Cham
www.phdcomics.com

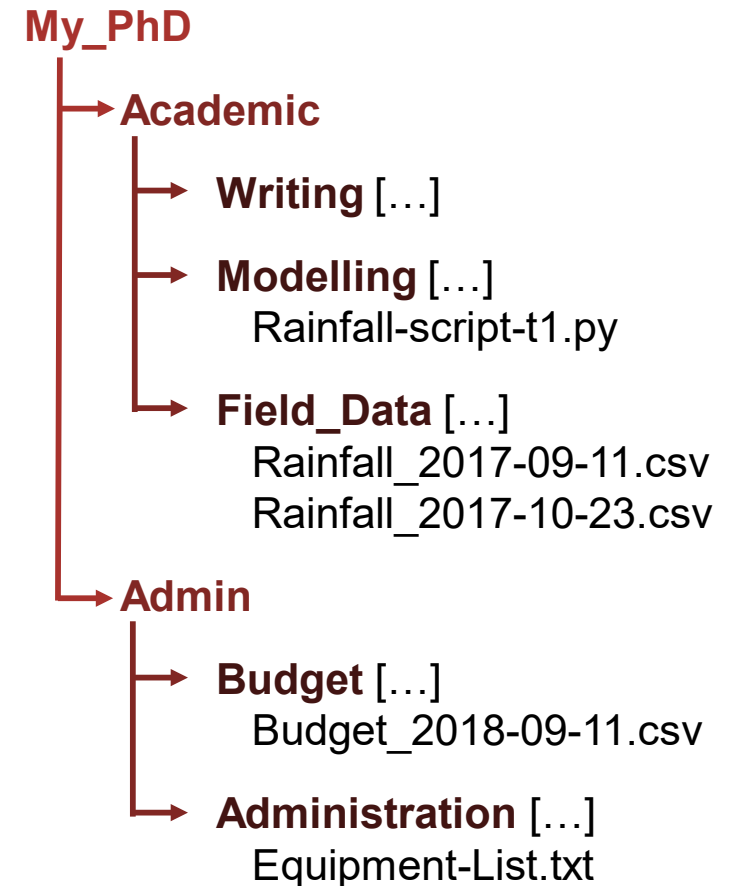
Versuchen Sie es lieber so:

- Dateien nach **Themen** sortieren
- Ordnerpfade **kurz** halten
- **Eindeutige** und möglichst **kurze** Dateinamen
- **Keine Sonderzeichen**
- **nicht** von **Case Sensitivity** ausgehen (name = NaMe)
- Datum nach ISO: **YYYY-MM-DD**
- Organisationsregeln in einem **Readme festhalten**

www.data.cam.ac.uk/data-management-guide/organising-your-data

datalib.edina.ac.uk/mantra/organisingdata/

www.wur.nl/en/Expertise-Services/Data-Management-Support-Hub/Browse-by-Subject/Organising-files-and-folders.htm



Strategien für Forschungsgruppen



■ **Selbstkritische Fragen:**

- Wie müssen Daten aussehen, um sie mit wissenschaftlicher Überzeugung und Vertrauen in Qualität und Korrektheit wiederverwenden zu können?
- Stimmt das für unsere eigenen Daten? Was fehlt?

■ **Aufgaben für Gruppenleiterinnen und -leiter**

- Vereinbaren Sie verbindliche Regeln
- Bestimmen Sie einen Datenmanagement-Verantwortlichen innerhalb der Gruppe
- Besprechen und dokumentieren Sie Regeln (schriftlich) mit dem Datenmanagement-Verantwortlichen



Blick in den Werkzeugkasten

Gruppendiskussion zur aktuellen Nutzung

- **Versionierung:** Wie gehen Sie damit um? Was funktioniert gut? Was läuft schief?
- **Namensregeln:** Benutzen Sie sie und falls ja, welche?
- **Teilen:** Welche Werkzeuge und Dienste verwenden Sie? Was sind Ihre Erfahrungen?
- **Literaturmanagement:** Welche Programme verwenden Sie? Was sind ihre Vor- und Nachteile?
- **Sonstige?** z.B. fachspezifische Plattformen oder zentrale Datenbanken?

Kriterien zur Auswahl von Dienstleistungen und Tools

- **Standort** der Dienstleister und ihrer Server
- Rechtsvorschriften zum **Datenschutz**
- **Nachhaltigkeit** und **Vertrauenswürdigkeit**
- **Zugriffsrechte** auf Ihre Daten
- **Lizenzen** und unmittelbare/längerfristige **Kosten**
- Wie bekommen Sie Ihre **Daten zurück?**



Repositorien und Registries

zenodo.org

www.re3data.org

datadryad.org

www.openaire.eu/search/data-providers

figshare.com*

*nur bedingt empfehlenswert, Daten können laut Nutzungsbedingungen jederzeit ohne Ankündigung seitens figshare gelöscht werden



Zusammenarbeit – Teilen / Sharing

Empfohlen

- Daten in der Schweiz
- Sicherheitsbestimmungen erfüllt

polybox.ethz.ch

cifex.ethz.ch/

www.switch.ch/drive/

www.switch.ch/filesender



SWITCH



Nur bedingt empfohlen

- Daten in EU/USA
- Sicherheitsbestimmungen nur z.T. erfüllt
- Nie vertrauliche / Private Daten dort ablegen

www.dropbox.com

www.wetransfer.com

Zusammenarbeit – Organisation / Versionierung

Organisation

www.openproject.org

asana.com

slack.com

www.redmine.org

tagpacker.com

trello.com

Versionierung

ETH Zürich: [Sharepoint](#)

subversion.apache.org

github.com

bitbucket.org



Zusammenarbeit – Schreiben / Referenzen

Schreiben

ETH Zürich: [Sharepoint](#)

[www.overleaf.com](#)

[hypothes.is](#)

[www.authorea.com](#)

[evernote.com](#)

[www.onenote.com](#)

[atlas.oreilly.com](#)

[simplenote.com](#)

Referenzmanagement

[www.mendeley.com](#)

[endnote.com](#)

[www.zotero.org](#)

[www.jabref.org](#)

[www.citeulike.org](#)

[www.bibsonomy.org](#)



Laborinformationsmanagementsysteme / Elektronisches Laborbuch

Kommerziell

- an der ETH Zürich in Verwendung

benchling.com

findingsapp.com

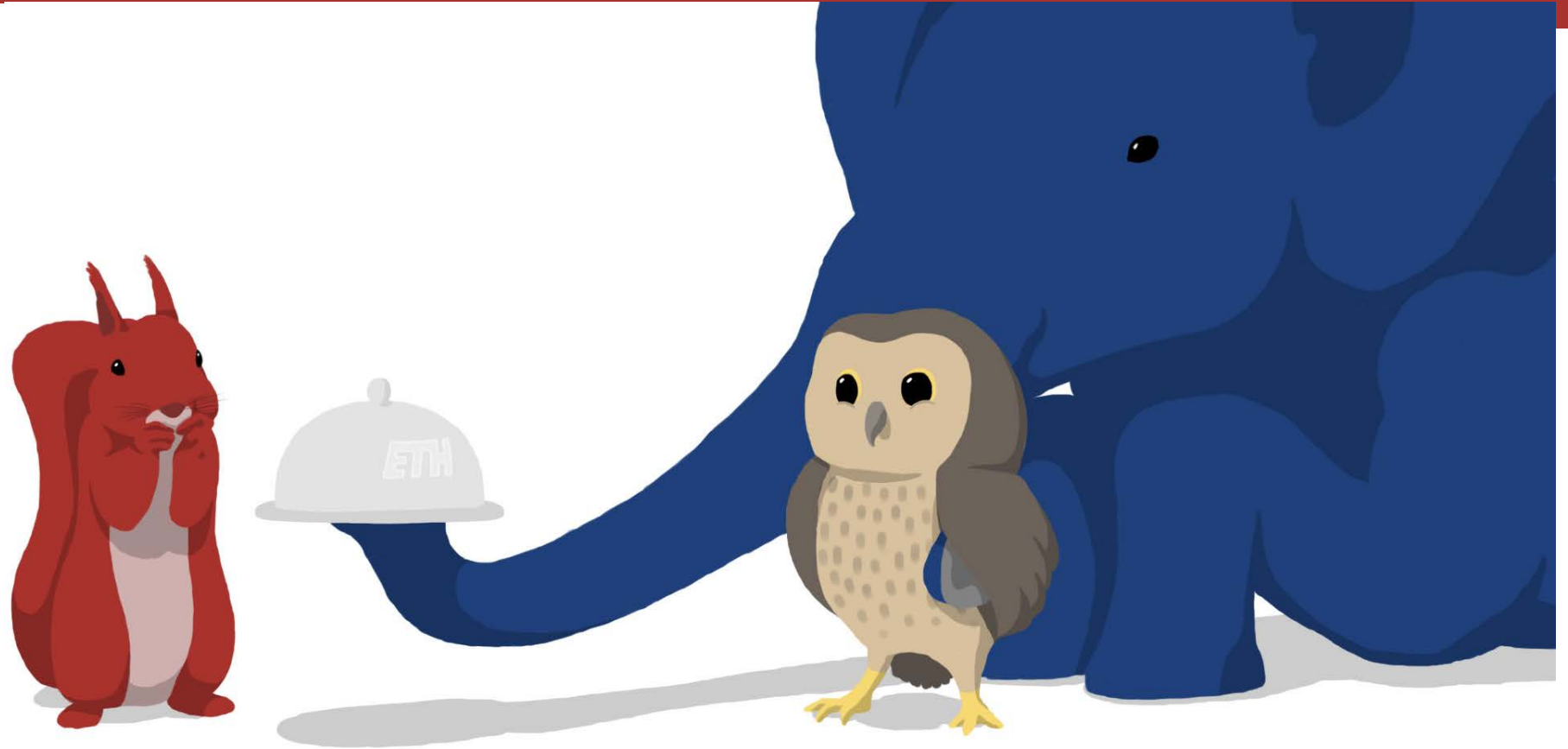
labcollector.com

Hausintern

- Anpassung möglich
- Datenspeicherung an der ETH Zürich
- Speichert Proben, Protokolle, Daten, Resultate, Analysen/Scripts und Beschreibungen der Experimente

<https://labnotebook.ch/>





Dienstleistungen der ETH Zürich

Dienstleistungen der ETH-Bibliothek

ETH-Bibliothek

- Research Collection:
www.research-collection.ethz.ch
- ETH Data-Archive:
www.library.ethz.ch/Digitaler-Datenerhalt
- DOI Registrierung:
www.library.ethz.ch/DOI-Desk
- Open Access:
www.library.ethz.ch/de/Open-Access
- ORCID*: www.library.ethz.ch/ORCID

*Verknüpfung mit Ihrer ETH-Identität möglich

Schulungen

- Kurse und Workshops zu Datenmanagement, Literaturverwaltung, Open Access, Informationssuche, Wissenschaftlichem Schreiben, „Book a Librarian“
www.library.ethz.ch/de/Dienstleistungen/Schulungen-Fuehrungen
- Forschungsdatenmanagement Workshops in Zusammenarbeit mit ETH SIS
www.ethz.ch/researchdata
- Weitere Themen auf Anfrage
researchdata@ethz.ch

Dienstleistungen der ETH Zürich Informatikdienste

Datenspeicherung und Versionierung

- Speicher (via internes IT-Support-Team)
 - NAS (Network Attached Storage):
www.ethz.ch/services/de/it-services/katalog/speicher/nas.html
 - LTS (Long-Term Storage):
www.ethz.ch/services/de/it-services/katalog/speicher/lts.html
- Gitlab (hosting durch Informatikdienste):
gitlab.ethz.ch
- SharePoint (bis zu 1 GB gratis):
mysite.sp.ethz.ch

Aktives Datenmanagement während des Forschungsprozesses

- Support für aktives Datenmanagement:
www.sis.id.ethz.ch/researchdatamanagement
- openBIS
Laborinformationsmanagementsystem und Elektronisches Laborbuch:
<https://labnotebook.ch/>
- Entwicklung von massgeschneiderten Lösungen

Weitere Dienstleistungen der ETH Zürich

ETH transfer

- Beratung zu geistigem Eigentum, Patenten, Softwarelizenzierung etc.
- Software-Offenlegungs Workflow mit dem ETH Data Archive

www.transfer.ethz.ch

Statistik Beratung @ D-MATH

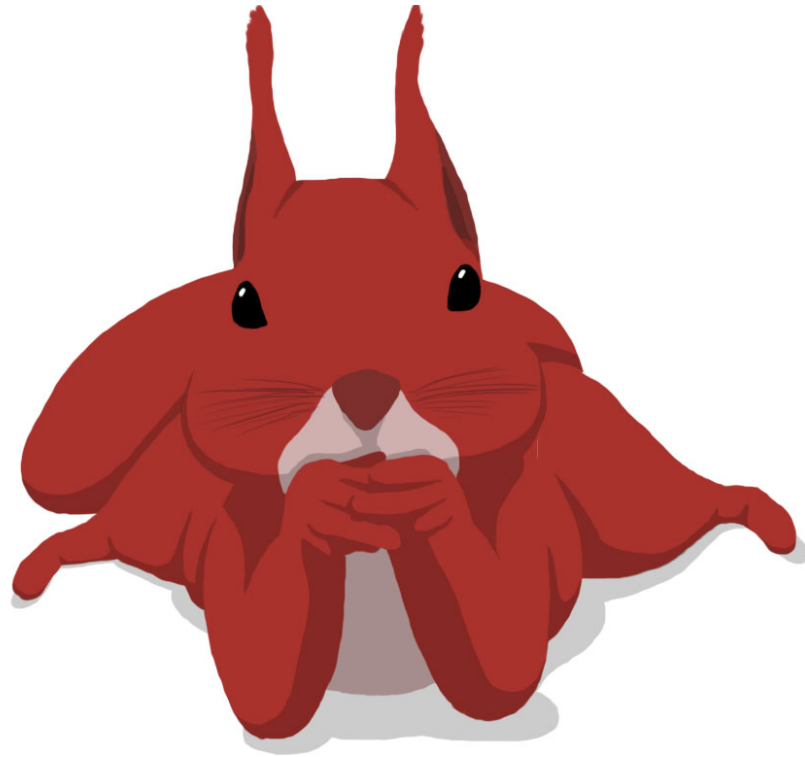
- Beratungsservice
- Statistik- und Softwarekurse

www.math.ethz.ch/sfs/consulting.html

Take home message

- **Überlegen** Sie, was Sie tun
- Beginnen Sie **frühzeitig**
- Treffen Sie **Vereinbarungen** über klare Konzepte und einfache Werkzeuge
- Sie brauchen **nicht immer raffinierte Apps**
- **Sprechen** Sie mit Kolleginnen und Kollegen
- Informieren Sie sich über das **Angebot Ihrer lokalen Dienstleister**
- «Halten Sie es so einfach wie möglich – aber mit gesundem Misstrauen!»

Wir benötigen Ihr Feedback



Bitte füllen Sie das Kursbewertungsformular aus – Vielen Dank!

<https://www.umfrageonline.ch/s/960fb89>

Danke!

Dr. Ana Sesartic
Digitaler Datenerhalt
ETH-Bibliothek
Rämistrasse 101
8092 Zürich
044 632 73 76
ana.sesartic@library.ethz.ch



www.library.ethz.ch/Digitaler-Datenerhalt
data-archive@library.ethz.ch