

Forschungsdatenmanagement - was kann ich tun?

Educational Material

Author(s):

Petrus, Ana 

Publication date:

2017-11-28

Permanent link:

<https://doi.org/10.3929/ethz-b-000297432>

Rights / license:

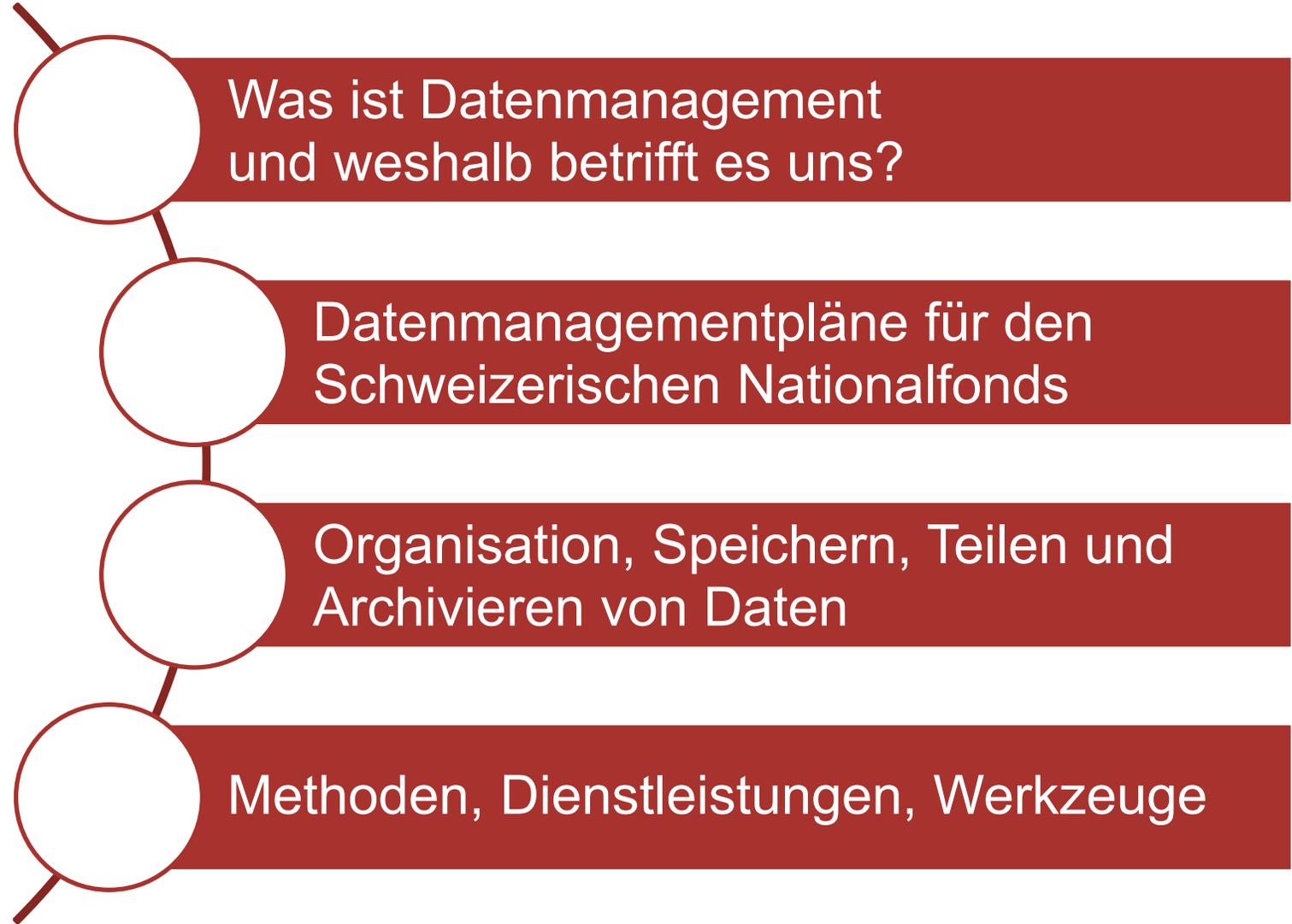
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International



Forschungsdatenmanagement – Was kann ich tun?

Dr. Ana Sesartic – Digitaler Datenerhalt

Heutiges Programm



Erzählen Sie kurz über...

- Ihren (wissenschaftlichen) **Hintergrund**
- Bisherige **Erfahrungen** mit Datenmanagement
- **Motivation** zur Kursteilnahme



Was ist Datenmanagement und weshalb betrifft es uns?

Kern des Forschungsdatenmanagements

«...nachvollziehen was man vor [einigen] Jahren gemacht hat, es wiederfinden und [...] unmittelbar wiederverwenden können.»

Henry Rzepa, Professor of Computational Chemistry, Imperial College London

Definition des Datenmanagements

- **Datenmanagement** ist ein allgemeiner Begriff, welcher die **Organisation, Strukturierung, Speicherung und Pflege** von Informationen beinhaltet, die während eines Forschungsvorhabens verwendet oder erzeugt werden
- Es umfasst:
 - **Wie Sie im Alltag** mit Informationen über die Laufzeit eines Projektes **umgehen**
 - **Was auf längere Sicht passiert** – was Sie damit machen, nachdem das Projekt abgeschlossen ist

Was sind Daten?



“Eine reinterpretierbare Repräsentation von Information in formaler Weise, brauchbar für Kommunikation, Interpretation und Prozessierung.”

© Digital Curation Centre
 Slide adapted from the PrePARe Project / [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/)

Einschränkungen für deren Erhalt und Austausch

**GARBAGE IN,
GARBAGE OUT!**



["MGB Grau Blau WP"](#) (4.9.2018) by [Bidgee](#) / [CC BY-SA 3.0](#)

- Daten werden in der Regel **erstellt, ohne** deren **Veröffentlichung zu planen**
- **Forschungsdaten bedürfen einer umfassenden Dokumentation**
- Nur technische Metadaten können später extrahiert werden, aber **wenig bis gar keine Dokumentation von Inhalt oder Kontext kann sinnvoll hinzugefügt werden.**

Weshalb Zeit und Mühe aufwenden?

- Sicherung nicht replizierbarer Daten
- Vermeidung redundanter Datenerhebungen
- Möglichst umfassende (kosteneffiziente) Auswertung der Daten
- Reanalyse unter anderen Fragestellungen / mit neuen Methoden
- Verstärkung des wissenschaftlichen **Austausches**
- Impact factor: Trend zur **Zitierbarkeit von Forschungsdaten** als Werk
- Um die **Vorgaben** der Geld- und Arbeitgeber zu **erfüllen**
 - SNF fordert Datenmanagementpläne ab Oktober 2017
 - Horizon 2020 Programm der EU fordert ebenfalls Datenmanagementpläne
- Einhaltung der Regeln **guter wissenschaftlicher Praxis**, Transparenz und validität
- Sie können die **Diskussion** in Ihrer Forschungsgemeinde, in Ihrer Institution und mit Geldgebern **beeinflussen**



© Seppo Leinonen: "Sticks and carrots"
<http://www.seppo.net/cartoons/displayimage.php?album=6&pid=1188> (4.9.2018)



“Preservation plan” by [Jørgen Stamp](#) / [CC BY 2.5](#)

Datenmanagementpläne für den Schweizerischen Nationalfonds

Was ist ein Datenmanagementplan (DMP)?

Ein kurzer Plan, der zu Beginn eines Projekts verfasst und während des Projektverlaufs aktualisiert wird, um zu definieren:

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

DMPs werden
verlangt vom:

SNF seit Oktober 2017

http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/default.aspx

Horizon 2020

EU funding programme

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Weshalb ein DMP?

DMPs werden zunehmend für Projektanträge verlangt, sind aber auch sonst nützlich, wann immer Forschende Daten erstellen.

Sie helfen Forschenden:

- **Bewusste Entscheidungen zu treffen**, um Probleme zu antizipieren und zu vermeiden
- **Konsistenzwahrende Prozeduren frühzeitig zu entwickeln**
- **Sicherzustellen, dass Daten korrekt, vollständig, zuverlässig und sicher sind**
- **(Unerwünschte) Duplizierung, Datenverlust und Sicherheitsverletzungen zu vermeiden**
- **Zeit und Mühe zu sparen und ihr Leben zu erleichtern!**

Datenmanagementpläne (DMP) gemäss SNF

- Ziel: **Forschungsdatenlebenszyklus planen und dokumentieren** von der Datenerstellung bis zum –Erhalt
- Daten **FAIR** machen: Findable – Accessible – Interoperable – Re-usable
- Zumindest die **Daten, welche Publikationen unterliegen, zugänglich** machen
- DMP muss ins **mySNF Webformular** eingetragen werden, zusammen mit dem Antrag
- Es ist ein **lebendiges Dokument**, das aktualisiert werden muss, bis das Projekt abgeschlossen wird
- **Endversion** des DMP wird in **P³** Projektdatenbank veröffentlicht
- **DMP ist nicht Teil der Gesuchsevaluation**, wird aber vom SNF nach seiner **Plausibilität** geprüft
- **Zusätzliche Finanzierung** (bis zu CHF 10'000.-) für Kosten der Datenaufbereitung zur Veröffentlichung
- Ratschlag vom SNF: falls ein Punkt nicht erfüllt werden kann, unbedingt kommentieren
- **Siehe Dokumentation vom SNF** unter http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx, und kontaktiere ord@snf.ch bei Fragen

Wie reicht man einen DMP bei SNF ein

- Ein Forschungsgesuch kann erst dann eingereicht werden, wenn auch der DMP ausgefüllt wurde.
- Der DMP für den SNF muss online auf der mySNF Plattform erstellt werden.
- DMP Inhalt:
http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_de.pdf
- Instruktionen zum Ausfüllen des SNF DMP für ETH – Forschende:
<http://www.library.ethz.ch/en/Media/Files/DL/CM-template-for-the-SNSF-Data-Management-Plan>

ETH SNF SCHWEIZERISCHER NATIONALFONDS ZUR FÖRDERUNG DER WISSENSCHAFTLICHEN FORSCHUNG mySNF

D F E

Login
Registrierung
Zugangsdaten vergessen

Willkommen bei mySNF

mySNF ist die Web-Plattform des Schweizerischen Nationalfonds zur Interaktion zwischen Gesuchstellenden, Evaluierenden, Forschungsinstitutionen und der Geschäftsstelle des SNF. Sie ermöglicht die Einreichung, Evaluierung und Verwaltung von Gesuchen und Projekten.

Bitte informieren Sie sich auf der Webseite des SNF (<http://www.snf.ch/>) über die verschiedenen Förderungsinstrumente. Eine Liste der nächsten Eingabetermine finden Sie unter [Ausschreibungen - SNF](#).

Um mySNF nutzen zu können, benötigen Sie ein Benutzerkonto. Unter "Neue/n Benutzer/in registrieren" können Sie sich für ein Benutzerkonto registrieren lassen.

Wichtig: Neue Benutzerkonten können erst nach einer Prüfung durch den SNF genutzt werden. Eine entsprechende Bestätigung über die erfolgreiche Erstellung des Benutzerkontos wird Ihnen anschliessend per E-Mail zugestellt.

Login

Benutzername [Benutzer/in neu registrieren](#)

Passwort [Zugangsdaten vergessen](#)

[Anmelden](#)

Weiterführende Links

- [Benutzerregistrierung und Anmeldung: Kurzanleitung & FAQ](#)
- [mySNF Benutzungsordnung](#)

<https://www.mysnf.ch>

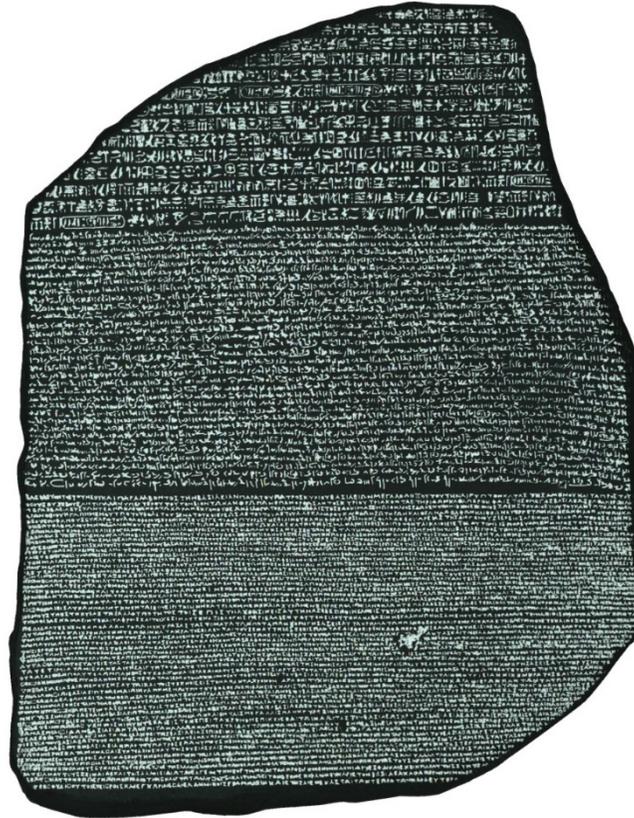
Was tun bei anderen Geldgebern?



- **Datenmanagement Checkliste der ETH / EPFL**
- Unterstützt Sie bei der Erstellung eines DMP oder bei der Diskussion von Datenmanagement im Allgemeinen, auch wenn sie kein DMP für Geldgeber erstellen müssen
- <http://bit.ly/rdmchecklist>
- **DMPOnline**
- Das DMPOnline-Tool des UK Digital Curation Centre hilft Ihnen, [Horizon 2020](#) konforme Datenmanagementpläne zu erstellen
- <https://dmponline.dcc.ac.uk>

Sammlung von DMP Beispielen:

<http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>



Langzeitarchivierung von Daten

Und wie man sich darauf vorbereiten kann



Was hat das mit Datenmanagement zu tun?

- Ein **ordnungsgemässes Datenmanagement** oder dessen Fehlen bestimmen, ob der Datenerhalt möglich ist
- Über eine **Zeitspanne von zehn Jahren** mag Datenmanagement allein genügen, es ist jedoch sinnvoll vorauszudenken
- Falls Daten über **längere Zeiträume** aufbewahrt und genutzt werden sollen, sind weitere Massnahmen nötig:
 - **Daten sollen so selbsterklärend wie möglich sein**, einschliesslich der Dokumentation über die verwendete Software oder noch besser, der Aufbewahrung der Software selber; einschliesslich z.B. Referenzoutputs von Modellalgorithmen
 - **Mehr Sorgfalt bei der Auswahl und Verwendung von Dateiformaten ist geboten**

Bevorzugte Eigenschaften von Fileformaten



- **Offene Standards** (nicht proprietär)
 - Falls proprietär, nach Möglichkeit konvertieren oder, falls nicht möglich, Programm zur Ansicht der Daten beifügen
- **Gut dokumentiert**
- **Weit verbreitet und unterstützt** durch viele Programme
- **Unkomprimiert** (oder zumindest verlustfrei komprimiert)
- **Unverschlüsselt**
- Im Zweifelsfall **Original behalten und Kopie erstellen** in einem offenen Format
- Sich nicht auf **Dateiendungen** verlassen
- Beachten, dass **Daten auf unterschiedlichen Betriebssystemen verwendet** werden können

Beispiele

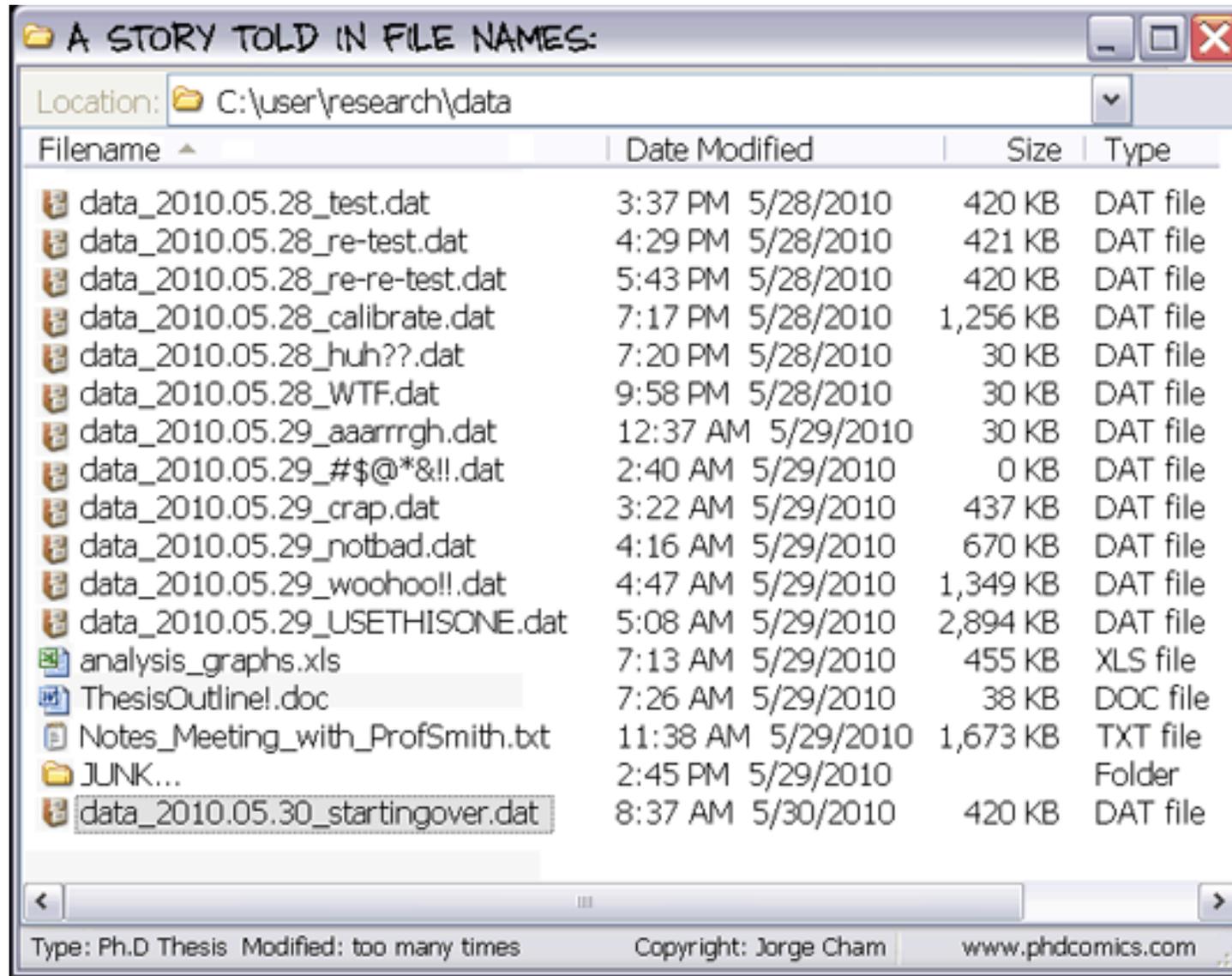
- **Bilder:** unkomprimierte TIFF; JPEG2000
- **Text:** ASCII, inklusive XML etc.
Informationen über **Kodierung** und **Abhängigkeiten** wie Stylesheets und TeX-Bibliotheken nicht vergessen
- **Text (formatiert):** PDF/A1-b, (PDF)
- **Daten aus Tabellen:** CSV
- **Tabellen:** (CSV), (ODF, OOXML)

Achtung

- **Dies bedeutet nicht, dass Sie die Daten nicht in anderen Formaten abspeichern dürfen**
- Sie müssen sich nur dessen bewusst sein, dass **proprietäre oder undokumentierte Formate** (auch Ihre eigenen!) **in der Zukunft Probleme verursachen können**
- Überlegen Sie sich **alternative Formate** (ja, redundant!) **zu den proprietären Formaten zu benutzen...**
- **...und Kontext-Informationen in einer Readme-Datei, Begleitdokument oder Metadaten hinzuzufügen** die sie selber in einigen Jahren gerne sehen würden, um die eigenen Daten verstehen zu können.

Organisieren Sie Ihre Daten

Finden Sie was Sie brauchen, wenn Sie es brauchen?



Kommt das Ihnen bekannt vor?

"A story told in file names" from
"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com

Source:

<http://www.phdcomics.com/comics/archive.php?comicid=1323>

Versuchen Sie es lieber so...

- Dateien nach Themen ordnen
- Klare und konsistente Benennungsregeln erstellen, die ihren Zweck erfüllen
- Dokumentieren Sie die Struktur und die Benennungsregeln in einer README Textdatei

Weitere Informationen finden Sie unter:

- <http://www.data.cam.ac.uk/data-management-guide/organising-your-data>
- <http://www.wur.nl/en/Expertise-Services/Data-Management-Support-Hub/Browse-by-Subject/Organising-files-and-folders.htm>
- <http://datalib.edina.ac.uk/mantra/organisingdata/>

Einige wichtige Punkte

- Zusammen lassen was zusammen gehört
- Pfadnamen kurz halten
 - < 255 Zeichen
- Dateinamen sollten
 - Inhalt widerspiegeln und eindeutig sein
 - Nur ASCII Zeichensatz verwenden (ohne Sonderzeichen)
 - Keine Leerzeichen enthalten
 - In Kleinbuchstaben oder «CamelCase» sein
- Achtung! Nicht alle Systeme unterscheiden zw. Gross- und Kleinbuchstaben!
 - UNIX: immer, Win/Mac: meistens
 - Nimm an, dass **das**, **DAS** und **dAs** gleich sind

- Datum so schreiben: YYYY-MM-DD

PUBLIC SERVICE ANNOUNCEMENT:

OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

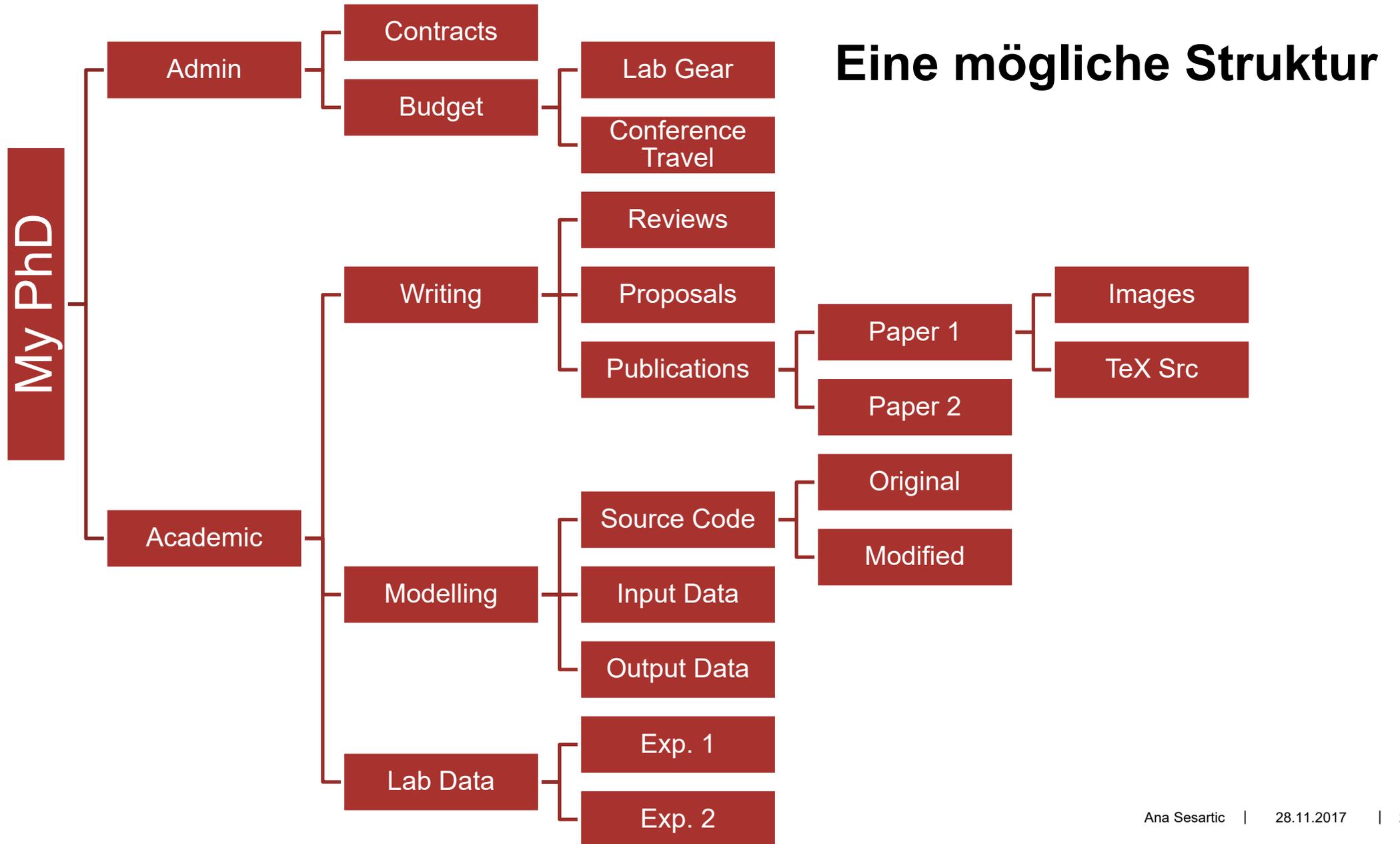
THIS IS *THE* CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
 20130227 2013.02.27 27.02.13 27-02-13
 27.2.13 2013.II.27. 27½-13 2013.158904109
 MMXIII-II-XXVII MMXIII ^{LVII}/_{CCCLXV} 1330300800
 ((3+3)×(111+1)-1)×3/3-1/3³ 2013 
 10/11011/1101 02/27/20/13 0²1³2⁴3⁷ 5 67 8

["ISO 8601"](#) (4.9.2018)
 by Randall Munroe
[CC BY-NC 2.5](#)



Was ist zu tun?

Strategien für Forschungsgruppen



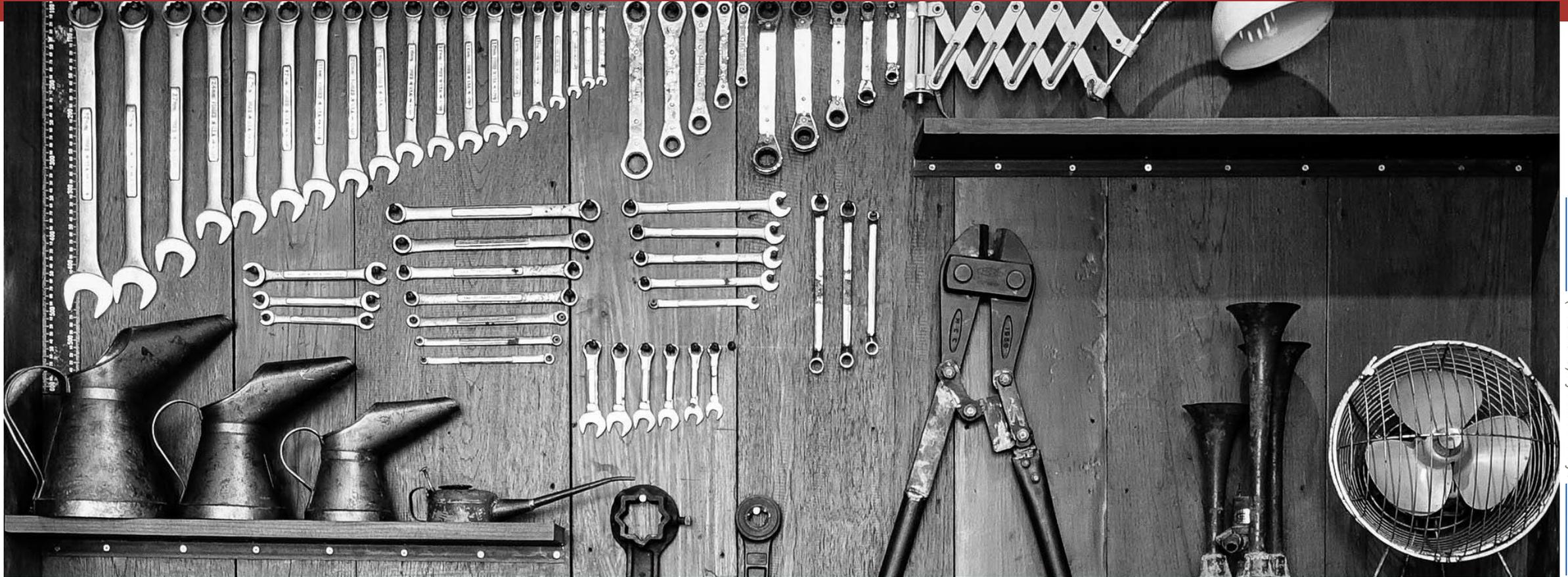
- **Selbstkritische Fragen:**
 - Wie müssen Daten aussehen, um sie mit wissenschaftlicher Überzeugung und Vertrauen in Qualität und Korrektheit wiederverwenden zu können?
 - Stimmt das für unsere eigenen Daten? Was fehlt?
- **Aufgaben für Gruppenleiterinnen und -leiter**
 - Vereinbaren Sie verbindliche Regeln
 - Bestimmen Sie einen Datenmanagement-Verantwortlichen innerhalb der Gruppe
 - Besprechen und dokumentieren Sie Regeln (schriftlich) mit dem Datenmanagement-Verantwortlichen

Rollen und Verantwortlichkeiten an der ETH



- Projekt **Mitglieder**:
 - Beachten die Grundsätze der guten wissenschaftlichen Praxis.
 - Sämtliche Verfahrensschritte im Umgang mit den Primärdaten sind derart zu **dokumentieren**, dass die gewonnenen Ergebnisse vollständig **reproduziert** werden können.
- Projekt **Leitung**:
 - Für das Management der Daten (Aufbewahrung, Datenzugang, Einhaltung des Datenschutzes, etc.) verantwortlich.
 - Sorgt dafür, dass alle Projektmitglieder sich der ETH Guidelines bewusst sind. Ensures that all research project participants are aware of the guidelines.
 - Entscheidet in Absprache mit der Professorin, welche Projektteilnehmer nach ihrem Ausscheiden aus dem Projektteam oder der ETH Zürich Zugang zu den Primärdaten bzw. Den Materialien behalten sollen.

Aus: **Richtlinien für Integrität in der Forschung**, <https://doi.org/10.3929/ethz-b-000179298>



"About the tools" (24.9.2018) by Ann Heathly / CC BY-NC-ND

Werkzeuge

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

- **Wo** befinden sich Ihre Daten?
- Welche **Rechtsvorschriften** gelten, z.B. in Bezug auf Datenschutz?
- Ist der Dienst **nachhaltig**?
- **Vertrauen** Sie dem Anbieter?
- **Wer kann** auf welche Ihrer Daten **zugreifen** und diese nutzen?
- Wie bekommen Sie Ihre **Daten zurück**?
- Ist eine bestimmte **Lizenz** erforderlich?
- Gibt es unmittelbare oder längerfristige **Kosten**?



“What is DP” by [Jørgen Stamp](#) / [CC BY 2.5](#)

Beispiel: Zusammenarbeit und Teilen

Empfohlen

- Daten in der Schweiz
- Sicherheitsbestimmungen erfüllt



<https://polybox.ethz.ch>



<https://cifex.ethz.ch/>

SWITCH

<https://www.switch.ch/drive/>

<https://www.switch.ch/filesender>

Nur bedingt empfohlen

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



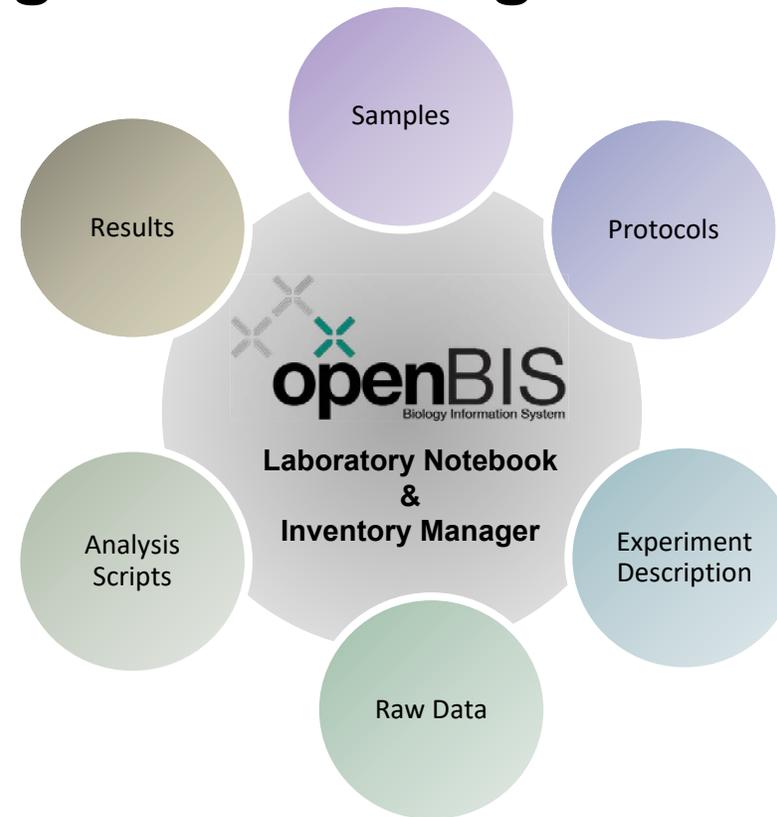
<https://www.dropbox.com>

wetransfer

<https://www.wetransfer.com>

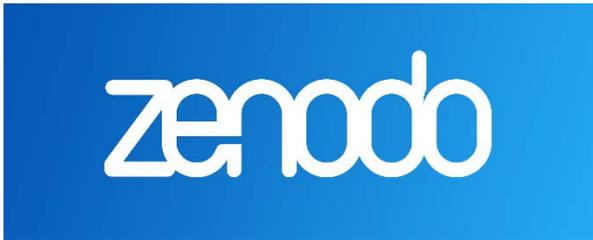
Hausinterne Lösung für Forschungsdatenmanagement / ELN

- Anpassung möglich
- Datenspeicherung an der ETH
- Speichert:
 - Proben
 - Protokolle
 - Beschreibung der Experimente
 - Daten



<https://labnotebook.ch/>

Repositorien und Registries



<https://zenodo.org>



<http://datadryad.org>



<http://figshare.com>

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



OpenAIRE

<https://www.openaire.eu/search/data-providers>

Zusammenarbeit - Organisation



<https://www.openproject.org>



<https://asana.com>



<https://slack.com>



<http://www.redmine.org>



<https://tagpacker.com>



<https://trello.com>

Zusammenarbeit - Versionierung



<https://subversion.apache.org>



<https://github.com>



<https://www.ethz.ch/services/en/it-services/catalogue/web-application-hosting/sharepoint.html>



<https://bitbucket.org>

(Sharepoint ist nur für die Versionierung von Dokumenten gedacht, nicht für Forschungsdaten!)

Zusammenarbeit - Schreiben



<https://www.overleaf.com>



<https://www.ethz.ch/services/en/it-services/catalogue/web-application-hosting/sharepoint.html>



hypothes.is

<https://hypothes.is>



<https://www.authorea.com>



<https://evernote.com>



<https://www.onenote.com>

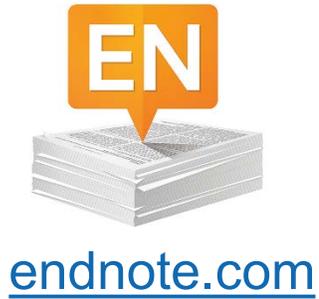


<https://atlas.oreilly.com>



<http://simplenote.com>

Zusammenarbeit – Referenzmanagement





Weitere Dienstleistungen und Weiterbildungen

Informatikdienste und ETH transfer

Informatikdienste

- Speicher (meist via internes IT-Support-Team)
- Support für Active Research Data Management www.sis.id.ethz.ch/researchdatamanagement
- openBIS **E**lectronic **L**ab **N**otebook & **L**aboratory **I**nformation **M**anagement **S**ystem
<https://labnotebook.ch/>

Versionierung

- Gitlab - gitlab.ethz.ch (Hosting durch IT services)
- SharePoint - mysite.sp.ethz.ch (gratis bis zu 1 GB)

ETH Transfer <https://www.ethz.ch/en/the-eth-zurich/organisation/staff-units/eth-transfer.html>

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

Schulungen

- Schulungen zur Recherche, Literaturmanagement, Datenmanagement, wissenschaftlichem Schreiben und Open Access bei der ETH-Bibliothek:
<http://www.library.ethz.ch/en/Services/Training-courses-guided-tours>
- Umfangreicher Workshop zu Datenmanagement offeriert von der ETH-Bibliothek in Zusammenarbeit mit den Scientific IT Services:
siehe oberen Link oder fragen Sie nach zusätzlichen Daten!
- Kurse des ETH Informationszentrums Chemie/Biologie/Pharmazie:
<http://www.infozentrum.ethz.ch/en/whats-up/events/>
- Weitere Themen nach Anfrage

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
- Informieren Sie sich über das Angebot Ihrer lokalen Dienstleister
- «Halten Sie es so einfach wie möglich – aber mit gesundem Misstrauen!»

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

Wir benötigen Ihr Feedback



Bitte füllen Sie das Kursbewertungsformular aus – Vielen Dank!