

Train-the-Trainer Workshop zum Thema Forschungsdatenmanagement

Tag 1

Katarzyna Biernacka, Petra Buchholz, Sarah Ann Danker,
Dr. Dominika Dolzycka, Claudia Engelhardt, Kerstin Helbig, Dr. Juliane Jacob,
Dr. Janna Neumann, Dr. Carolin Odebrecht, Dr. Ute Trautwein-Bruns,
Cord Wiljes, Dr. Ulrike Wuttke



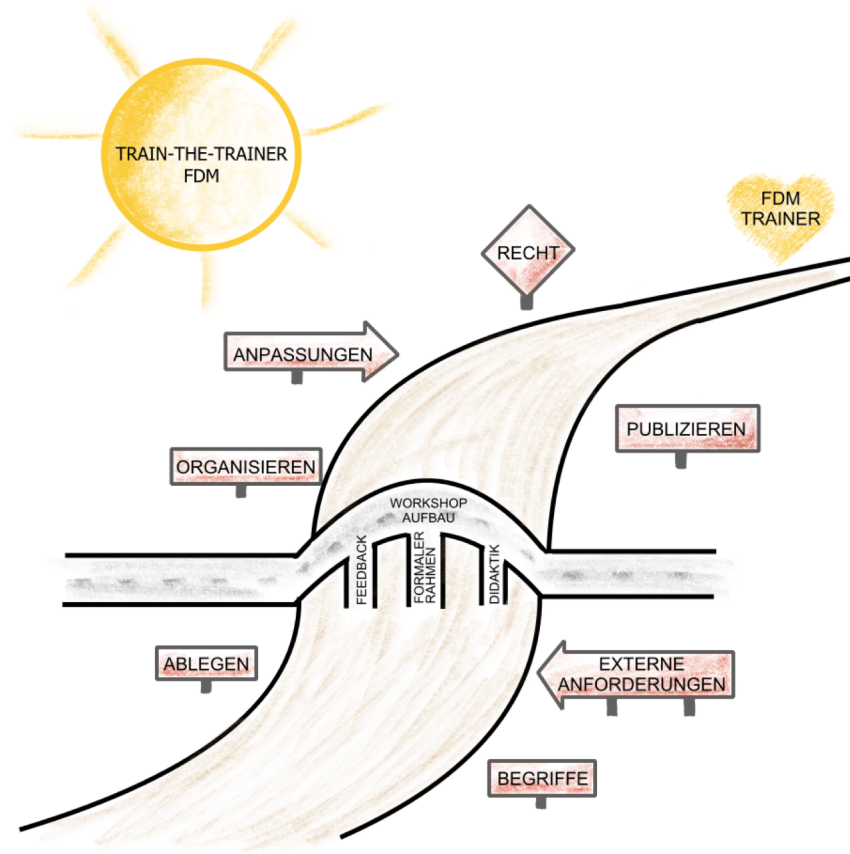
Begrüßung und Kennenlernen



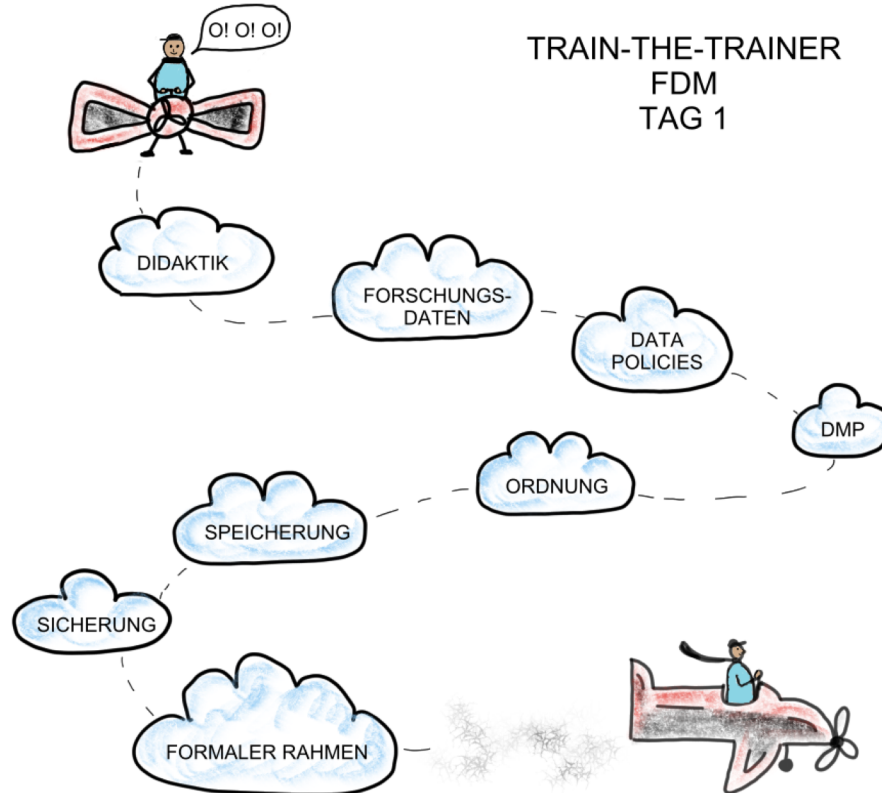
O! O! O! – Orientierung



Workshoplandkarte

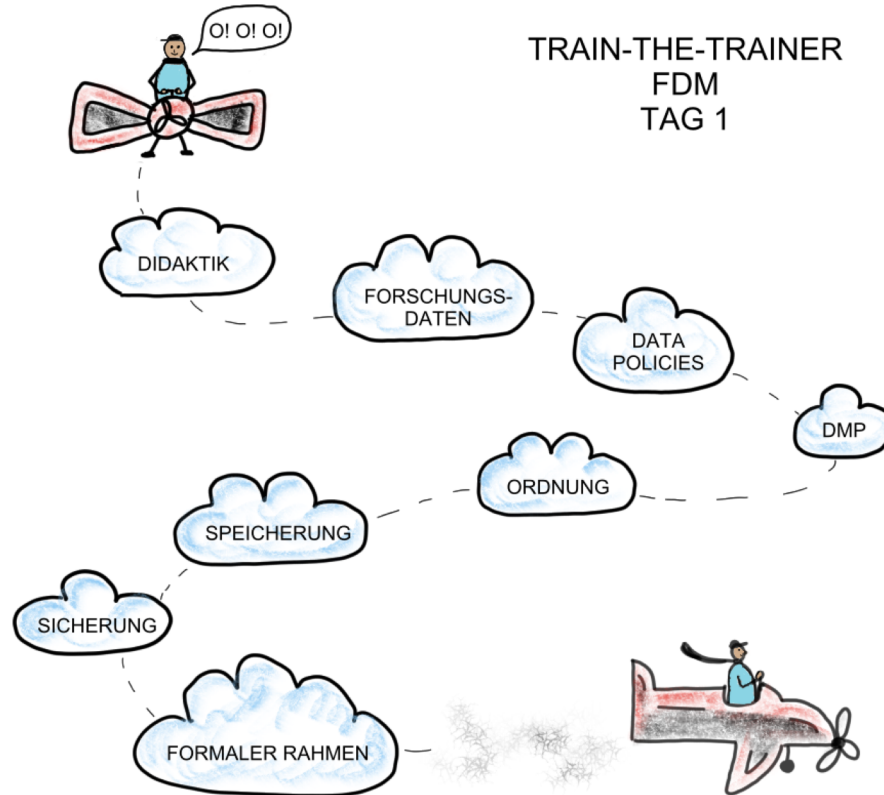


Tagesplan



Tagesplan

9:00 Beginn
~10:30 Kaffeepause
~12:00 Mittagspause
~14:30 Kaffeepause
16:30 Abschluss



Didaktisches Vorgehen



Übung

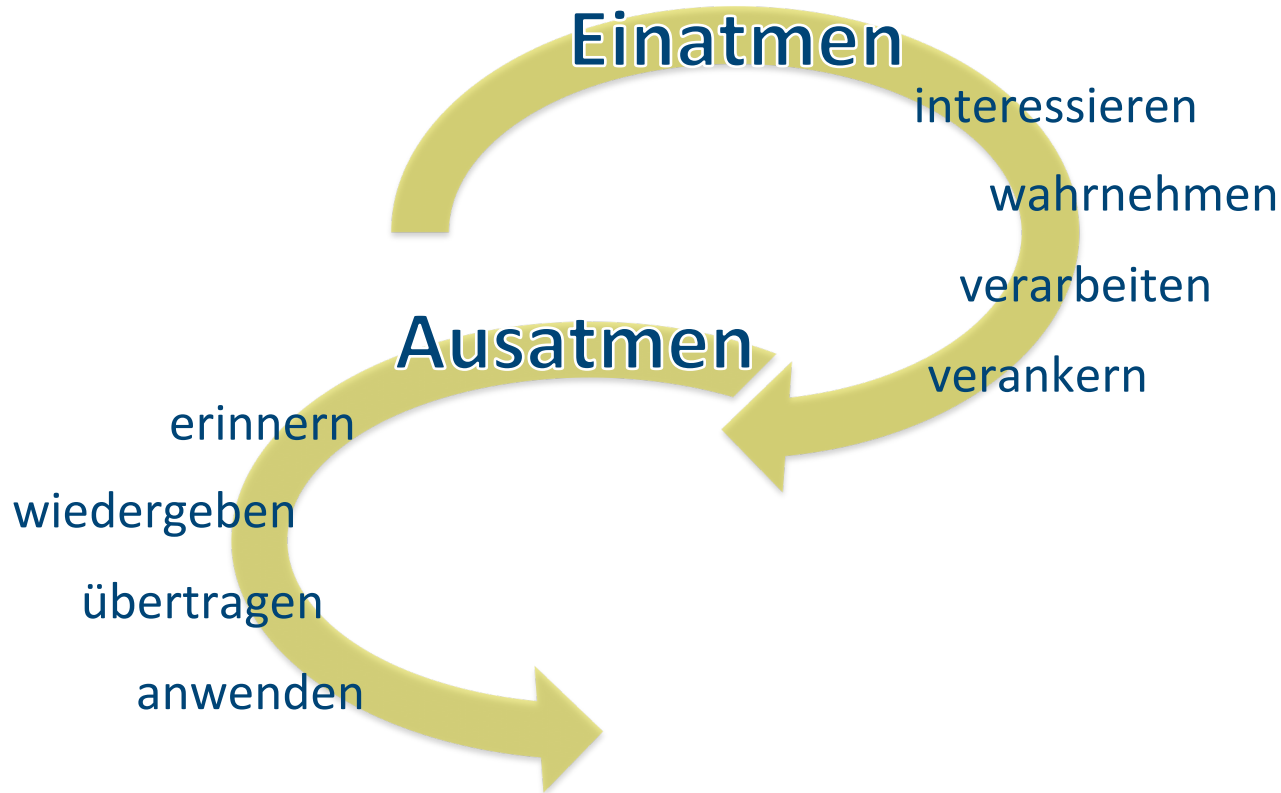
Welche Kriterien unterscheiden einen **Train-the-Trainer-Workshop** zum Thema Forschungsdatenmanagement von einem **Forschungsdatenmanagement-Workshop**?



Kriterien eines Train-the-Trainer-Workshops



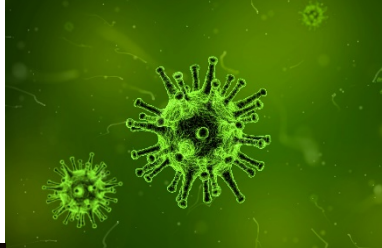
Lernprozess nach Klaus Döring



Digitale Forschungsdaten



Mit welchen Forschungsdaten arbeiten Sie?



Quelle: pixabay.com.
Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public
Domain Dedication.](https://creativecommons.org/publicdomain/zero/1.0/deed.de)
[https://creativecommons.org/publicdo
main/zero/1.0/deed.de](https://creativecommons.org/publicdomain/zero/1.0/deed.de)



Was sind digitale Forschungsdaten?

- Keine feste Definition von Forschungsdaten
- Allgemein:

„alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind“¹

[1]: Maxi Kindling, Peter Schirmbacher: *Die digitale Forschungswelt als Gegenstand der Forschung*. IWP. 2013. pp. 127-136. DOI: 10.1515/iwp-2013-0017.

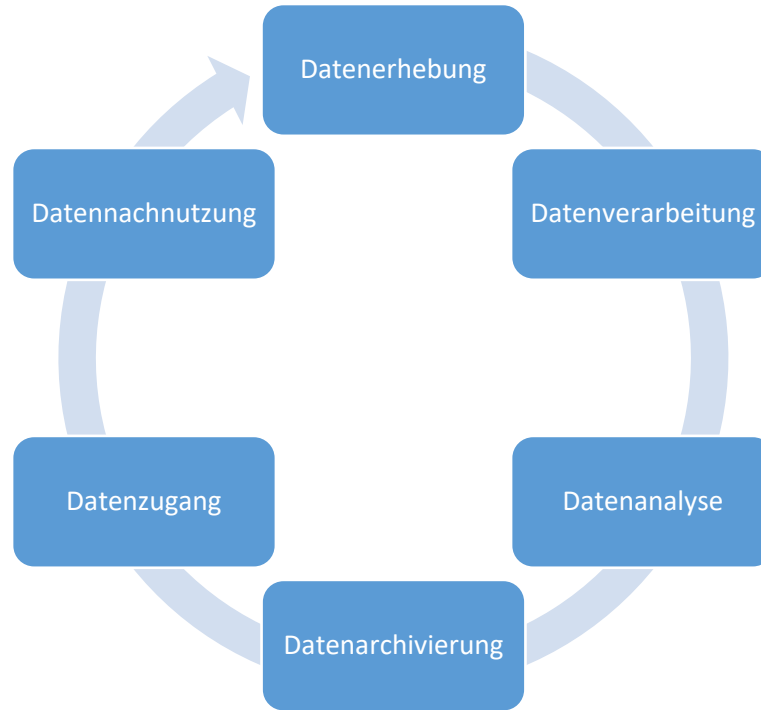


Übung

Forschungsdaten-Lebenszyklus



Auflösung: Forschungsdaten-Lebenszyklus



Quelle: Forschungsdaten-Lebenszyklus nach dem UK Data Archive (Stand: Juni 2018). Inzwischen ist dort eine neuere Version des Lebenszyklus dargestellt.



Forschungsdatenmanagement

Forschungsdatenmanagement (FDM) umfasst alle Aktivitäten, die mit

- Aufbereitung,
- Speicherung,
- Archivierung und
- Veröffentlichung

von Forschungsdaten verbunden sind.

FDM begleitet den Forschungsprozess von den ersten Planungen bis zur Archivierung, Nachnutzung oder Löschung der Daten.

Quelle: Maxi Kindling, Peter Schirmbacher, Elena Simukovic: *Forschungsdatenmanagement an Hochschulen: das Beispiel der Humboldt-Universität zu Berlin*. LIBREAS. Library Ideas, 23 (2013).
Online verfügbar unter: <http://libreas.eu/ausgabe23/07kindling/> DOI: 10.18452/9041

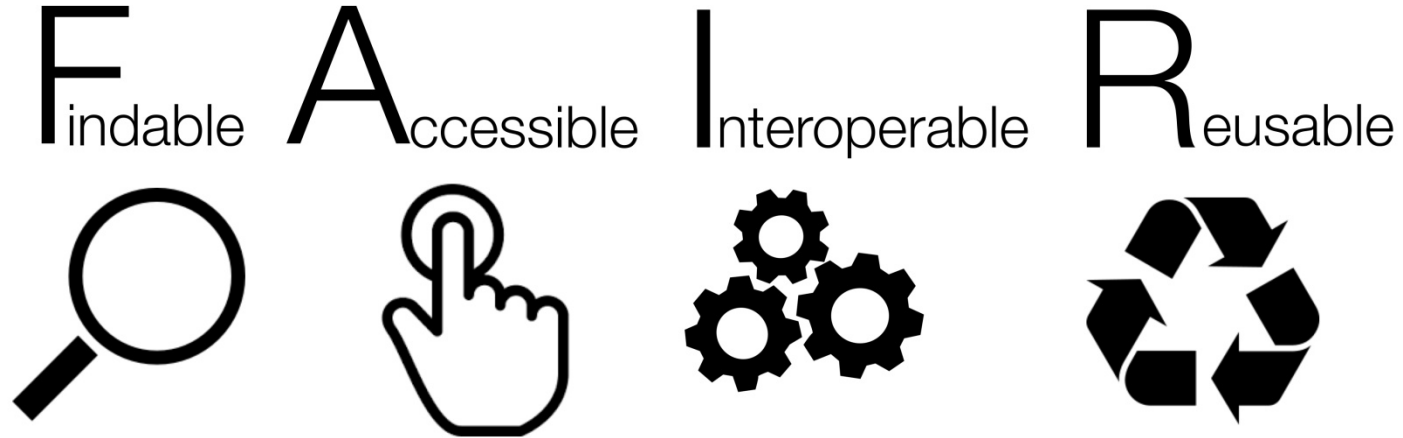


Übung

Welchen Nutzen bringt Forschungsdatenmanagement?



Die FAIR-Prinzipien



Quelle: Pundir, Sangya. https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg [letzter Zugriff: 27.09.2018], CC-BY-SA-4.0
<https://creativecommons.org/licenses/by-sa/4.0/deed.en>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand J. Aalberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>.



Aspekte des Forschungsdatenmanagements

- Ordnung und Strukturierung
- Dokumentation und Metadaten
- Speicherung und Backup
- Langzeitarchivierung
- Sicherheit
- Publikation
- Rechtliche Aspekte



Kaffeepause



Forschungsdaten-Policies



Forschungsdaten-Policy

Eine Forschungsdaten-Policy beschreibt die Richtlinien zum Umgang mit Forschungsdaten, z. B.:

- Zeitschriften- oder Verlags-Policies
- Förderrichtlinien
- Fachspezifische Policies
- Institutionelle Policies



Verlags-Policies - Beispiel Springer Nature

Typ 1 - Es wird dazu ermutigt, Daten zu teilen und zu zitieren
(Photosynthesis Research)

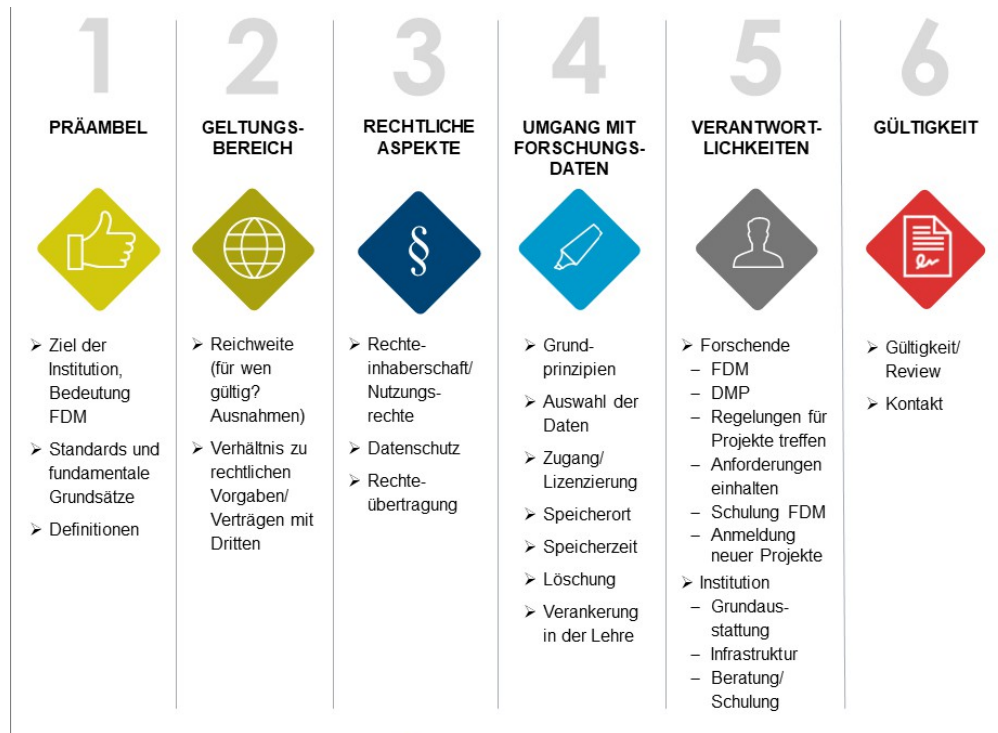
Typ 2 - Es wird dazu ermutigt, Daten zu teilen und
Datenzugänglichmachung nachzuweisen (Plant and Soil)

Typ 3 - Es wird ermutigt, Daten zu teilen und Aussagen zur
Datenverfügbarkeit sind verpflichtend (Palgrave Communications)

Typ 4 - Es ist eine Voraussetzung, Daten zu teilen, dies
nachzuweisen und ein Peer Review der Daten zu ermöglichen
(Scientific Data)



Inhalte einer institutionellen Forschungsdaten-Policy



Quelle: Hiemenz, Bea & Kuberek, Monika (2018). *Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland.* <http://dx.doi.org/10.14279/depositonce-7521>



Übung

Gibt es an Ihrer Einrichtung eine Forschungsdaten-Policy?
Welchen Umfang hat sie? Was wird wie geregelt?

und / oder

Würden Sie sich eine Forschungsdaten-Policy wünschen?
Welche Inhalte sollte sie haben?



Hilfestellung bei der Erstellung einer institutionellen Policy

Hiemenz, B.; Kuberek, M.: (2018). *Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland*. DOI: [10.14279/depositonce-7521](https://doi.org/10.14279/depositonce-7521).



Datenmanagementplan



Was sind Datenmanagementpläne?

- Alle Informationen, die die Sammlung, Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten im Rahmen eines Forschungsprojekts hinreichend beschreiben und dokumentieren
- „[...] Analyse des Workflows von der Erzeugung der Daten bis zu deren Nutzung“²

[2] J. Ludwig, H. Enke (Hrsg.) *Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt*. Verlag Werner Hülsbusch: Glückstadt, 2013. ISBN: 978-3-86488-032-2



Übung

Welche Vorteile kann ein Datenmanagementplan bringen?



Anforderungen der Forschungsförderer

Förderer	Forderung	Abgabe bei Antrag	Inhalt	Updates
EC Horizon 2020	DMP	Nein. Erster Plan innerhalb der ersten 6 Projektmonate	Inhalte des Horizon 2020 Template	Update, falls signifikante Änderungen auftreten sowie zum Projektende
DFG	Angaben zum Umgang mit Forschungsdaten	Ja	Inhalte der Leitlinie zum Umgang mit Forschungsdaten	Nein
BMBF	Plan manchmal erforderlich, abhängig vom Programm	Falls notwendig, ja	Inhalt hängt vom jeweiligen Programm ab; Bildungsforschung: Checkliste	Kommt auf das Programm an
VWStiftung	DMP	Ja	Inhalte des Science Europe Template	Nein



Bestandteile eines DMP

- Administrative Informationen (Projektname, Datenurheber, weitere Mitwirkende, Kontakt, Förderprogramm usw.)
- Projekt- und Datensatzbeschreibung
- Datentypen, -formate, -umfang
- Angaben zu Metadaten und Standards
- Daten teilen
- Archivierung und Sicherung der Daten
- Verantwortlichkeiten
- Kosten

Der Umfang kann zwischen wenigen Absätzen und mehreren Seiten variieren.



RDMO



Mit dem Research Data Management Organiser (RDMO) können Institutionen und Forschende das Forschungsdatenmanagement ihre Projekte strukturiert planen und durchführen. Es erlaubt das Erfassen aller relevanten Planungsinformationen in Datenmanagementplänen und die Verwaltung aller Datenmanagementaufgaben über den gesamten Datenlebenszyklus.

RDMO ist bereit für die Anwendung in kleineren und größeren Projekten. In der zweiten Phase des RDMO-Projekts seit November 2017 erweitern die Projektpartner AIP, FHP und KIT-Bibliothek die bereits veröffentlichte Version des RDMO und arbeiten vertieft mit Anwendern zusammen. Im Fokus der Erweiterung von RDMO stehen u.a. Rollenkonzepte und Anbindungen an die institutionelle Infrastruktur wie Repositorien, Ticket-Systeme und Infrastruktur für Authentifizierung und Autorisierung. Es werden Workshops zur Diskussion der Funktionalitäten durchgeführt, Einführungsmaterialien erstellt und eine Verbreiterung der Entwickler-Community angestrebt.

Auf diesen Seiten informieren wir über unser Projekt mit seinen Arbeitspaketen, die Software, deren Dokumentation, Aktuelles, unsere Veranstaltungen und unsere Kooperationspartner.

Die freie Software finden Sie unter:
github.com

Eine Demo-Instanz finden Sie unter:
rdmo.aip.de

Um sich für unsere Mailingliste anzumelden, klicken Sie auf folgenden Link:
rdmo@listserv.dfn.de

Folgen Sie uns auf Twitter:
twitter.com/rdmorganiser

Diskutieren Sie mit uns in Slack, einer webbasierten Software:
rdmo.slack.com

Wenn Sie Fragen haben oder uns anderweitig kontaktieren möchten, schicken Sie bitte eine Mail an unsere interne Mailingliste rdmo.team@listserv.dfn.de



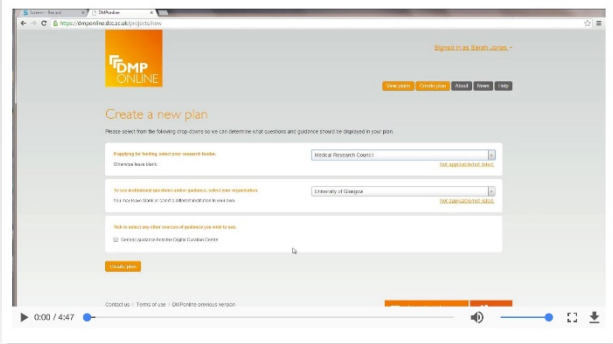
DMPonline



Welcome,

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

Screenshot on how to use DMPonline



Contact us | Terms of use

© 2004 - 2018 Digital Curation Centre (DCC)

Sign in

[Forgot your password?](#)

☐ Remember me

[Sign in](#)

Or, sign in with your [institutional credentials](#) (UK users only)

Create account

New to DMPonline? Create an account today.



Ordnung und Struktur



Übung

Wann habe ich schon mal Ordnung und/oder Struktur meiner Daten schmerzlich vermisst?

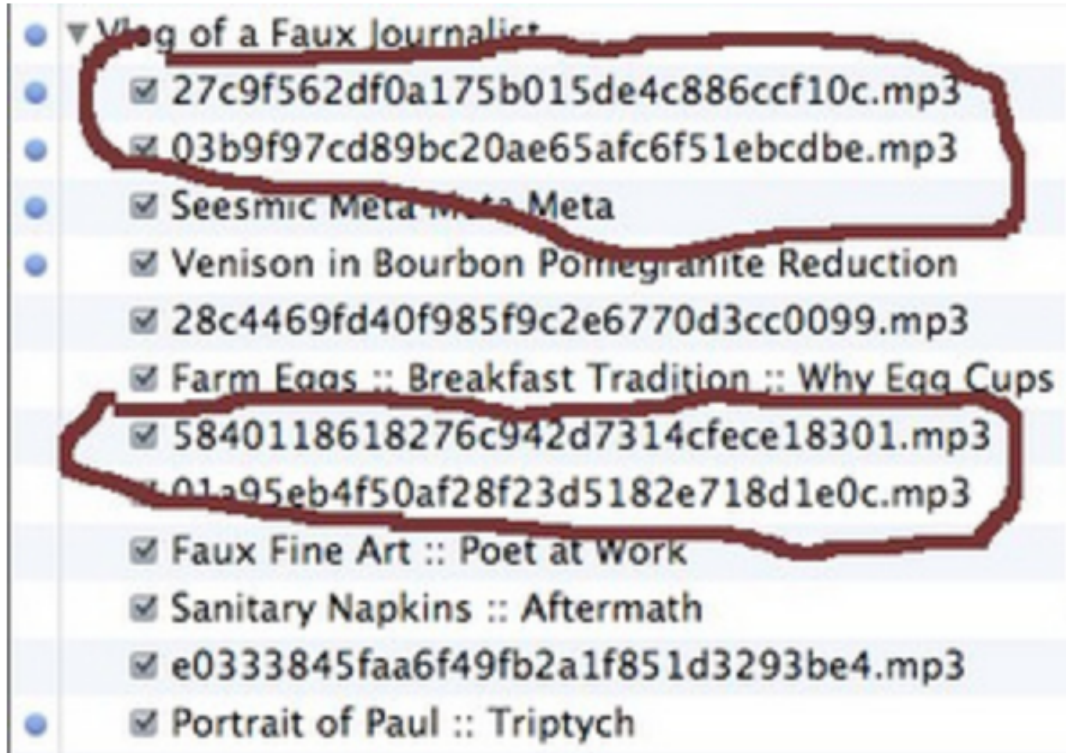


Wirkung von Struktur

- Was, wie und weshalb getan wurde, bleibt auch nach Jahren nachvollziehbar
 - Andere können mit den Daten auch arbeiten
 - Doppelte Arbeit wird vermieden
 - Datenverlust wird vermieden
- ➡ es kann effizienter gearbeitet werden



Benennung von Dateien und Ordnern



Quelle: MANTRA – Research Data Management Training:
Organising data.
<http://datalib.edina.ac.uk/mantra/organisingdata>, [Letzter
Zugriff: 01.03.2018]



Benennung von Dateien und Ordnern

- Aussagekräftige Namen -> keine „Phantasienamen“
- Einheitliches Schema
- Logische Struktur
- Datumsangabe zur chronologischen Sortierung in folgender Form: JJJJMMTT
- Vermeidung von Leer- und Sonderzeichen
- Dokumentierte Namenskonventionen oder genutzte Abkürzungen, z. B.

[Sediment]_[Probe]_[Instrument]_[JJJJMMTT].csv
[Projekt]_[Interview]_[Ort]_[Personen-ID]_[JJJJMMTT].mp4



Beispiele

- 20180312_h2oProbe1_original.jpg
- 20180315_h2oProbe1_KDS_Ausschnitt.jpg
- 20180324_h2oProbe1_KDS_Ausschnitt_bearbeitet_bunt.jpg



Übung: Welche dieser Beispiele folgen einer guten Benennungskonvention?

- Olga_170413_probe17k
- Naturepaper karl britta james fertig!
- Vm4520132Schmidt.pdf
- 647749157.pdf
- 170413_probe17k_olga
- Naturepaper+karl+britta+james &nal
- Olga170413probe17k
- Krst_765_spkt_1203
- Naturepaper+karl+britta+james fertig! überarbeitet
- Kristall_765_spektr_20161203
- Nature_karlbrittajames_endendversion
- 28q8QGfHKwrRw.pdf
- Tagung_Digitale_Wissenschaft.pdf



Übung

Bitte entwerfen Sie eine Benennungskonvention für Ihre Dateien und geben ein paar Beispiele.

Bitte entwerfen Sie eine Struktur für Ihre Ablage als Verzeichnisbaum.



Werkzeuge für gleichzeitige Umbenennung

Windows:

- Ant Renamer (www.anftp.be/software/renamer)
- RenameIT (sourceforge.net/projects/renameit)
- Bulk Rename Utility (www.bulkrenameutility.co.uk/)
- Total Commander
(<https://www.ghisler.com/deutsch.htm>)



Werkzeuge für gleichzeitige Umbenennung

Mac:

- Renamer 6 (for Mac) (renamer.com/)
- Name Changer (mrrsoftware.com/namechanger/)
- ExifRenamer




Werkzeuge für gleichzeitige Umbenennung

Beispiel ExifRenamer

ExifRenamer - Umbenennungs-Bestätigung

Alter Name: IMG_7112.jpg

Pre/Suffix: Wie in den Einstellungen festgelegt  Dateiname ist Ok!

Prefix:

Suffix:

☐ Dateivorschau anzeigen

☒ Datums-/Zeitkorrektur verwenden

Neuer Name:

Datums-/Zeitkorrektur

Datum/Zeit der Datei: 2016-06-16 12:00:46

Verwendete Zeitdifferenz: Keine Zeitdifferenz.

Jahr	Monat	Tag	Stunde	Minute	Sekunde
2016	06	16	12	00	46

Bitte geben Sie die Zeit im 24-Stunden Format an.
Sie wird dann entsprechend dem verwendeten
Dateinamensstil konvertiert.



Werkzeuge für gleichzeitige Umbenennung

Linux:

- GNOME Commander (www.nongnu.org/gcmd/)
- GPRename (<http://gprename.sourceforge.net/>)

Unix:

- **rename** command

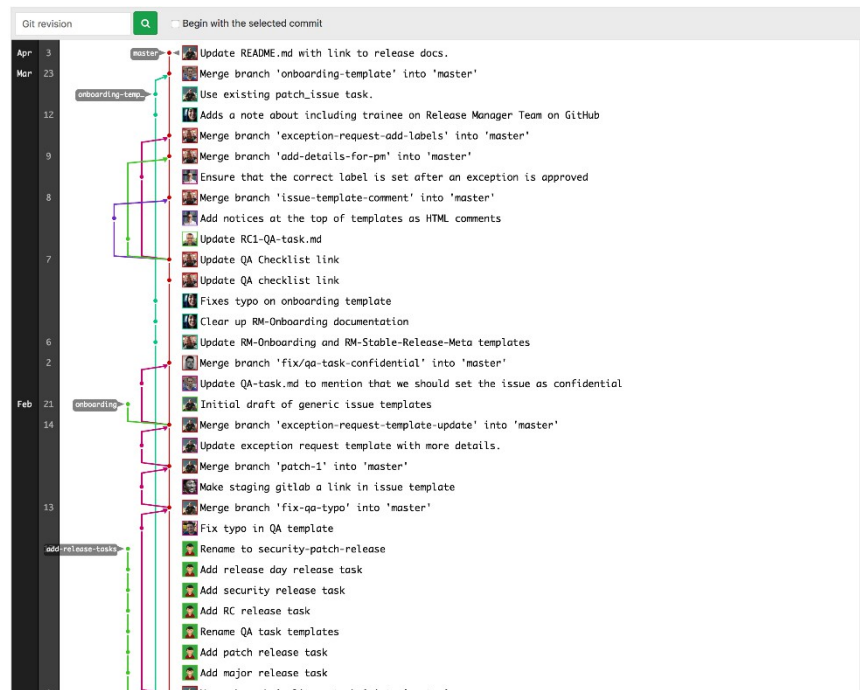


Kontrolle der Dateiversion

- obsolete Dateiversionen nach einem Backup separat, z. B. als eine separate ID (v1.0.0), ablegen
- Nutzung einer Versionskontrolltabelle
- Verantwortlichkeit für die Fertigstellung von Dateien festlegen
- bei großen Datenmengen ggf. Versionsverwaltungs-Software verwenden
- Meilenstein-Versionen speichern



Beispiel für Versionskontroll-Software (GitLab)



Mittagspause



Dokumentation und Metadaten



Übung

Warum ist eine Datendokumentation wichtig?



Übung

Schauen Sie sich die Tabelle an und nennen Sie Tipps, um die Datendokumentation zu verbessern.



Grundlegende Inhalte einer Dokumentation

- Beschreibung des Forschungsvorhabens
- Projektziele
- Hypothesen
- Informationen zur Erhebung der Daten (Methoden, Einheiten, Zeiträume, Orte, verwendete Technik)
- Maßnahmen zur Datenbereinigung
- Struktur der Daten und deren Beziehungen zueinander
- Erläuterung von Variablen, Labels und Codes
- Unterschiede zwischen verschiedenen Datensatz-Versionen
- Informationen zum Zugang und Nutzungsbedingungen



Metadaten – Daten über Daten

- Inhaltliche Metadaten:

- Titel
- Beschreibung
- Autor
- Urheberrechts-Inhaber
- Kontaktdaten
- Lizenzangaben
- Schlüsselwörter

- Technische Metadaten:

- Aufnahmedatum
- Brennweite
- Blende
- Belichtungsdauer
- Geographische Koordinaten
- und viele weitere



Normdaten

Dienen der eindeutigen Identifikation von Personen, Institutionen, Forschungsförderer...

- Gemeinsame Normdatei (GND)
- International Standard Name Identifier (ISNI)
- Virtual International Authority File (VIAF)
- Open Funder Registry



Kontrolliertes Vokabular

Klassifikation

dient der Zuordnung von Objekten in (meist hierarchisch strukturierten) Klassen. Die Klassen sind durch bestimmte Merkmale charakterisiert.

Thesaurus

Ist eine natürlich-sprachliche geordnete Sammlung von Begriffen und deren Beziehungen zueinander.



Übung

Erstellen Sie eine Liste von Metadaten, die Ihrer Meinung nach in Ihrem Fachgebiet vorkommen (können).



Beispiele für Metadatenstandards

Eine Übersicht zu disziplinspezifischen und fachübergreifenden Metadatenstandards gibt es unter:

<http://rd-alliance.github.io/metadata-directory/subjects/>

Beispiele:

- Fachübergreifende Metadatenstandards:
 - Dublin Core
 - MARC21
- Geowissenschaften:
 - ISO 19115
- Biodiversität:
 - Darwin Core
- Geisteswissenschaften:
 - Text Encoding Initiative (TEI)
- Naturwissenschaften:
 - ICAT Schema
 - Crystallographic Information Framework
- Sozialwissenschaften:
 - Data Documentation Initiative



Speicherung und Backup



Übung

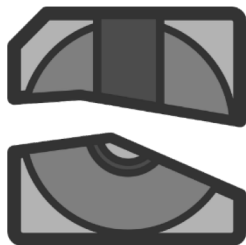
Bewerten Sie die verschiedenen Speichermedien.



Schutz vor Datenverlust

- Risiken

- Technische Defekte
- Katastrophen (Unwetter ...)
- Diebstahl
- Vergesslichkeit
- ...



- Strategien

- Speicherung auf Universitätsservern mit automatischem regelmäßigem Backup
- Sicherung wichtiger Dateien in mindestens drei Kopien auf räumlich getrennten Datenträgern



Quelle: pixabay.com.
Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public
Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/deed.de).
[https://creativecommons.org/publicdo
main/zero/1.0/deed.de](https://creativecommons.org/publicdomain/zero/1.0/deed.de)



Kriterien für Backup

Einrichtung von Backups

- Mindestens 3 Kopien einer Datei
- Auf mindestens 2 unterschiedlichen Medien
- Wovon mindestens eins dezentral ist

Testen Sie die Datenwiederherstellung zu Beginn sowie in regelmäßigen Abständen.

Schützen Sie Ihre (sensiblen) Daten:

- Hardware (bspw. separater abschließbarer Raum)
- Dateiverschlüsselung
- Sicherheit der Passwörter
- Mindestens zwei Personen sollten Zugang zu Ihren Daten haben



Kaffeepause



Langzeitarchivierung



Wie lange ist Langzeit?



CD: 5-10 Jahre



Festplatte: 3-10 Jahre



Diskette: 10-20 Jahre

Quelle: pixabay.com.
Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public
Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/deed.de).
[https://creativecommons.org/publicdo
main/zero/1.0/deed.de](https://creativecommons.org/publicdomain/zero/1.0/deed.de)



Archivierung – Abgrenzung zu Backup

Backup

Archivierung



Archivierung – Abgrenzung zu Backup

Backup

- (Automatische) Sicherung **aller** Daten, um Datenverlust vorzubeugen (technisch, z. B. defekt, oder menschlich, z. B. versehentlich gelöscht)
- Alle Versionen

Archivierung

- Sicherung **ausgewählter** Daten, um diese langfristig aufzubewahren
- Nur endgültige Versionen
- Integritätssicherung
- Langzeitspeicherung
- Durchsuchbarkeit



Nachhaltige Dateiformate

Dateiformat	Empfehlung	Vermeiden
Tabellen	CSV, TSV, SPSS portable, XLSX	XLS, SPSS
Text	TXT, HTML, RTF, PDF/A, DOCX	DOC, PDF
Multimedia	Container: MPEG4, MKV Codec: Theora, Dirac, FLAC	QuickTime, Flash
Bilder/Grafiken	TIFF, JPEG2000, PNG	GIF, JPEG



Übung

Erarbeiten Sie Kriterien für die Auswahl eines Langzeitarchivs:

Worauf sollte bei der Wahl geachtet werden?



Kriterien für die Auswahl von Langzeitarchiven



Zugriffssicherheit



Gründe für einen sicheren Umgang mit Daten

- Vor Datendiebstahl schützen
- Missbrauch der Daten verhindern
- Personen schützen (sensible Daten)



Sicherheit von Daten

Weshalb ist Datensicherheit für Sie wichtig?



Aspekte der Datensicherheit

- Vertraulichkeit
- Integrität
- Verfügbarkeit
- Kontrollierbarkeit



Physischer Schutz und Verschlüsselung

- Physischer Schutz vor Zugriff, z. B. abschließbarer Raum, Safe, Datentreuhänder etc.
- Automatische Verschlüsselungsoptionen, z. B. FileVault, Bitlocker, dm-crypt

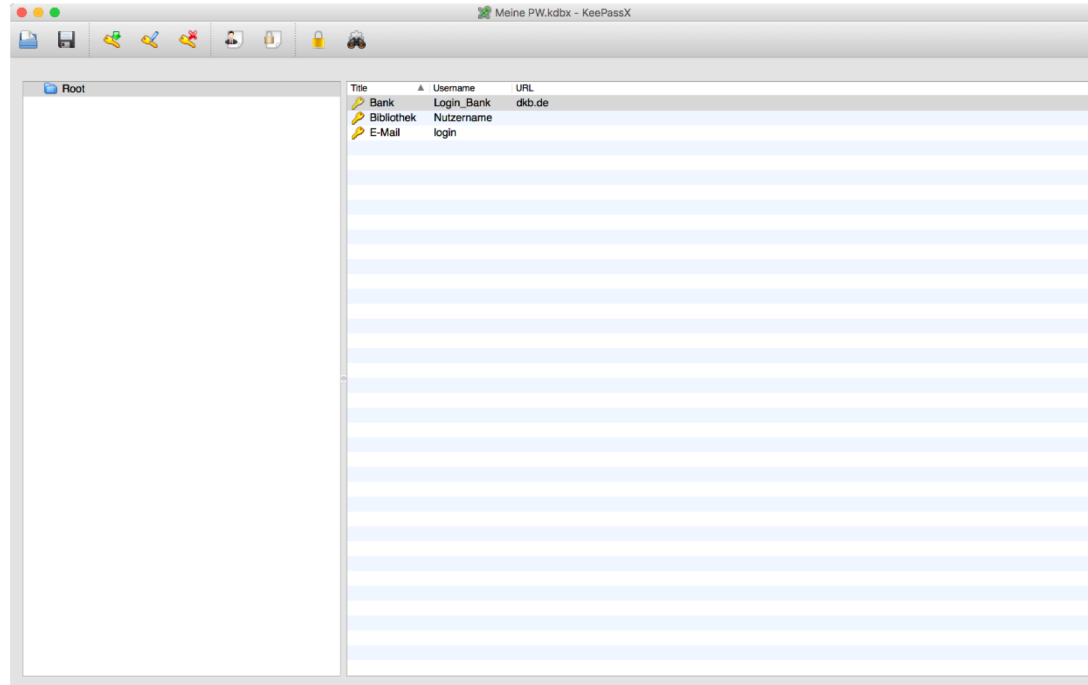


Passwortschutz

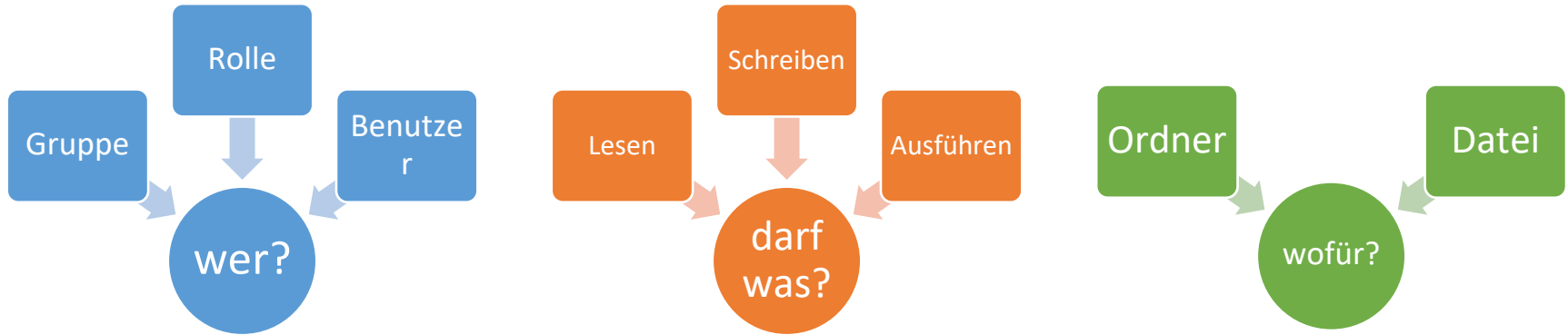
- Mindestens 8 Zeichen
- Klein- und Großbuchstaben
- Sonderzeichen und Zahlen
- Verwendete Zeichen sollten auf der Tastatur nicht nebeneinander liegen
- Nutzen Sie keine Passwörter, die in Wörterbüchern vorkommen



Beispiel: KeePassX



Zugriffsrechte und Rollenvergabe



Formaler Rahmen



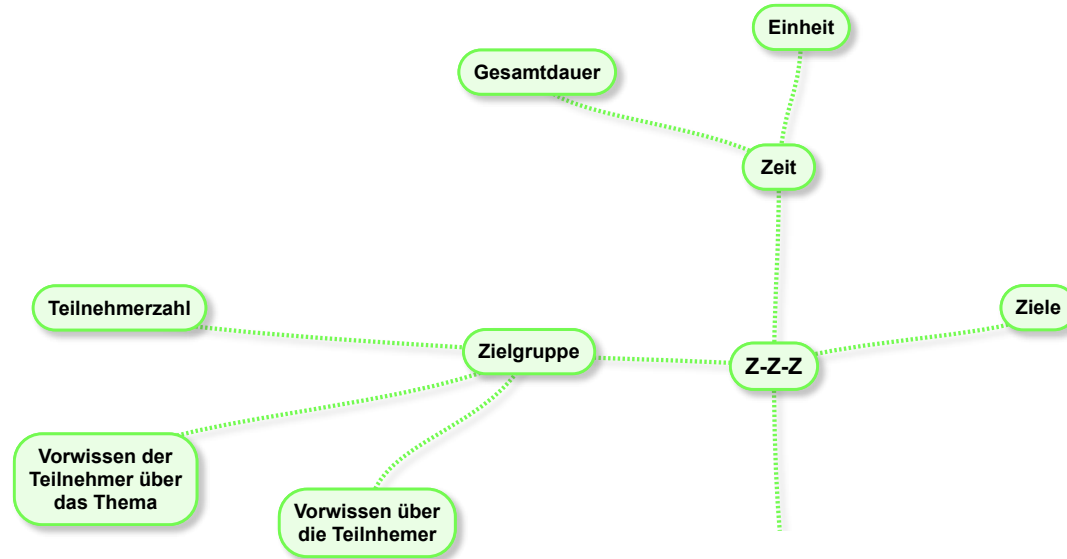
Übung

Gestalten Sie eine Mindmap zum Thema: Was muss ich bei der Planung eines Workshops bedenken?

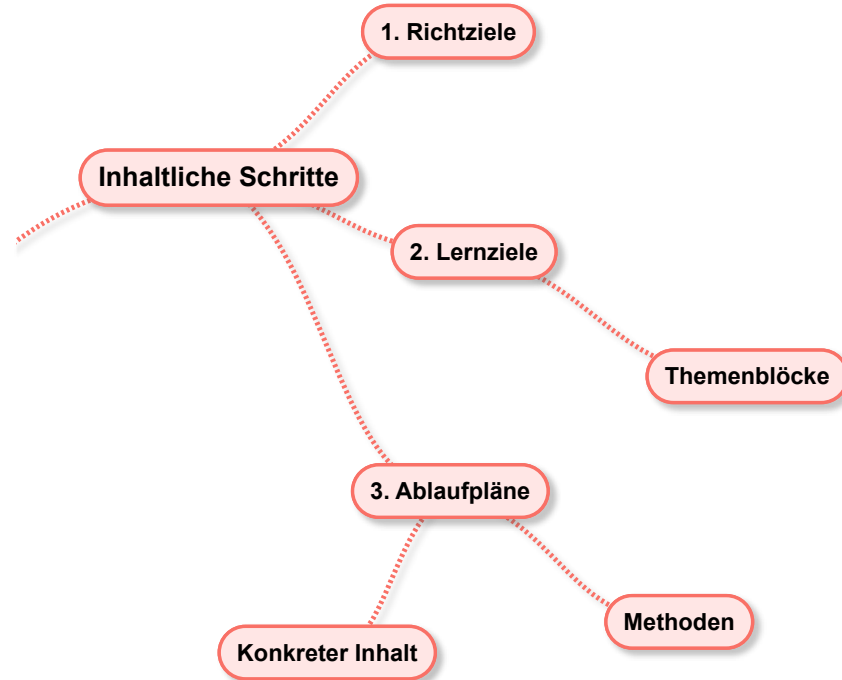
- Gruppe 1: Rahmenbedingungen
- Gruppe 2: Inhalte
- Gruppe 3: Organisation



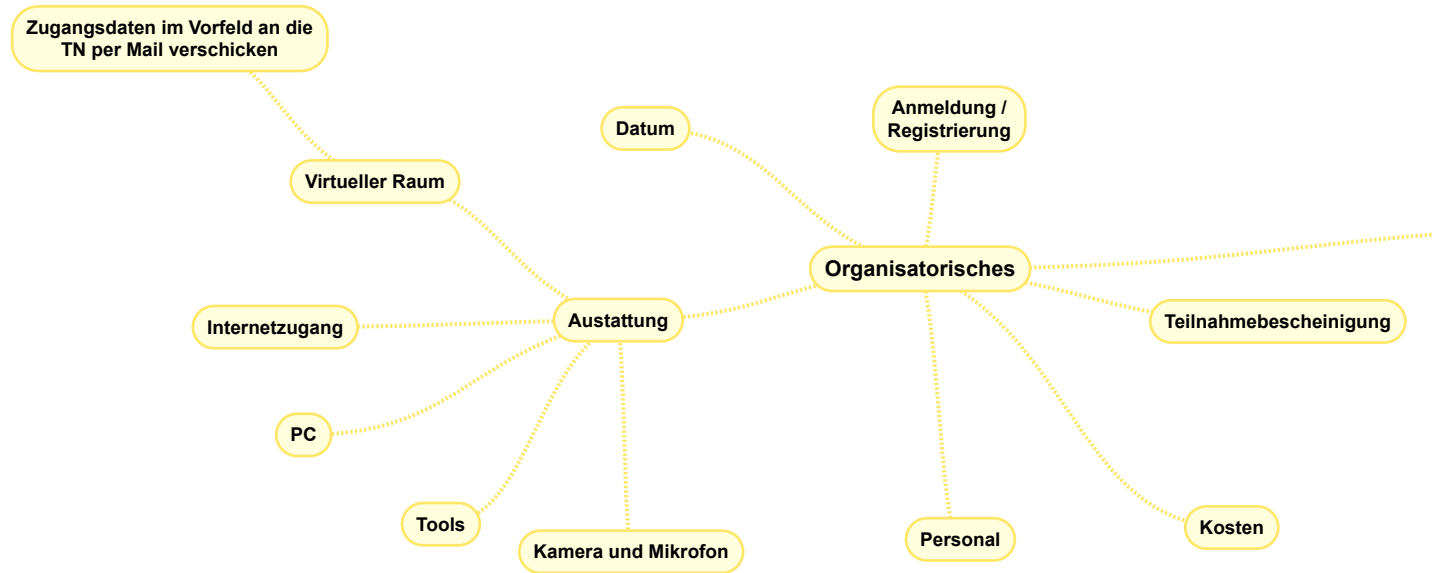
Erste Schritte: die 3Z-Formel

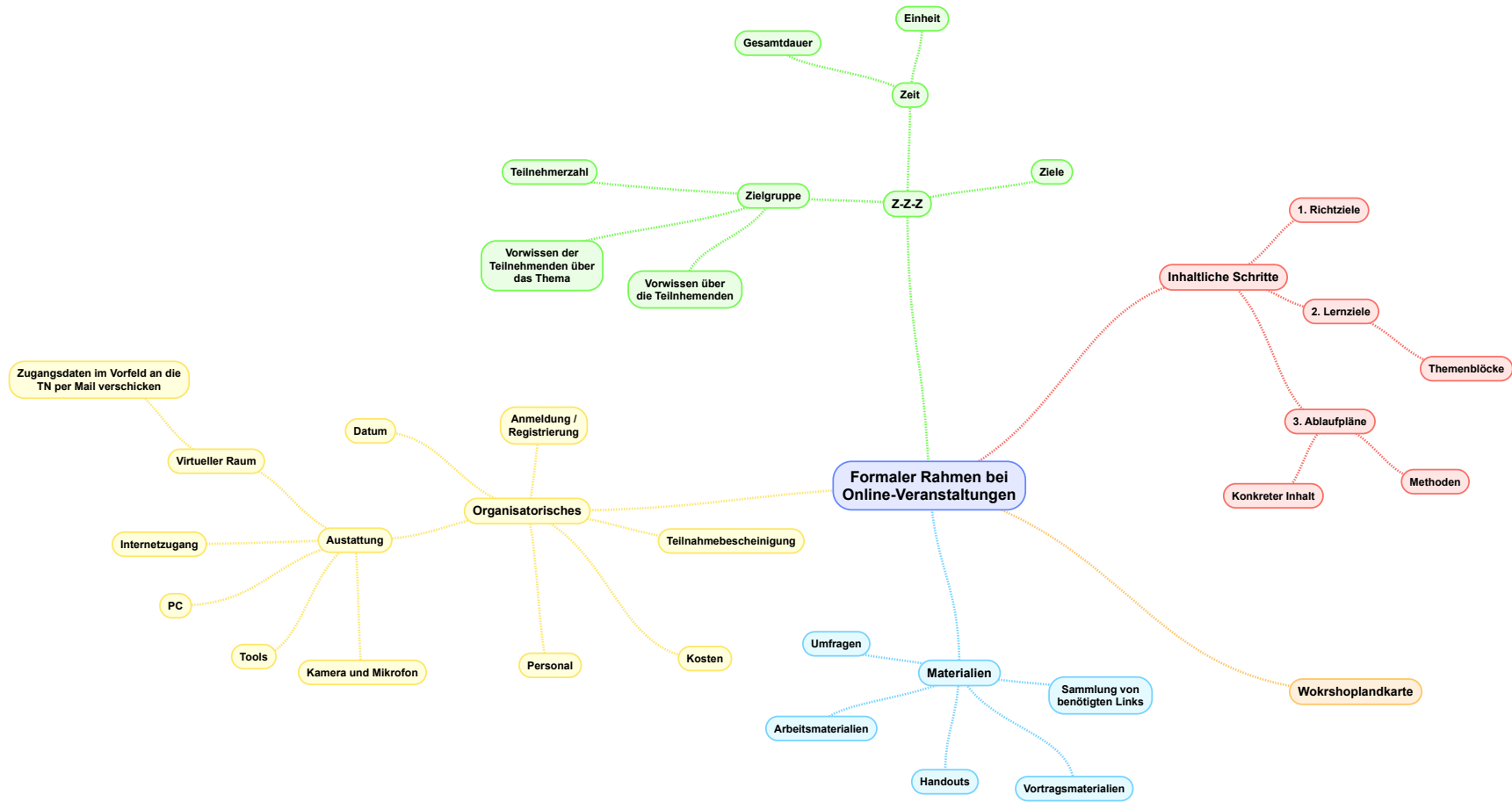


Inhaltliche Schritte



Organisatorisches





Übung

Was haben Sie heute gelernt?



Feedback

Stellen Sie sich vor, Sie schreiben einer/einem Freund*in oder Kolleg*in nach dem Workshop eine kurze Nachricht.

Wie würden Sie den heutigen Workshop in ein oder zwei Sätzen beschreiben?

Bitte schreiben Sie Ihre Nachricht in den Chat.



Quelle: pixabay.com.
Dieses Werk ist lizenziert unter einer
[CC0 1.0 Universal \(CC0 1.0\) Public
Domain Dedication](https://creativecommons.org/publicdomain/zero/1.0/deed.de).
[https://creativecommons.org/publicdo
main/zero/1.0/deed.de](https://creativecommons.org/publicdomain/zero/1.0/deed.de)



Fragen

Was möchten Sie uns noch sagen?



Herzlichen Dank für Ihre Teilnahme!

Katarzyna Biernacka
Petra Buchholz
Sarah Ann Danker
Dr. Dominika Dolzycka
Claudia Engelhardt
Kerstin Helbig

Dr. Juliane Jacob
Dr. Janna Neumann
Dr. Carolin Odebrecht
Dr. Ute Trautwein-Bruns
Cord Wiljes
Dr. Ulrike Wuttke

