

Datenmanagementplan – Fluch oder Segen?

Dr. Andreas von der Dunk
Team Forschungsdatenmanagement
SLUB Dresden

*„Forschende sollten in ihren Projekten
mit einem **DMP** arbeiten.“*

*Nicht **OBWOHL** sondern **WEIL**
ihr Forschungsalltag schon voll genug ist.“*

Teil 1

Daten und
ihr Wert

Teil 2

Der Nutzen
eines DMP
für die
Forschungsarbeit

Wert der Daten

“[Data] can be **bought, sold, shared** and even **stolen** – all things that suggest that data has **value**.”

“Yet ... very **few organizations** can attribute a value and, more concerning, many do not yet have the **capabilities to manage, protect** and **extract** that value.“

*Richard Petley
PwC Risk and Assurance*



Quelle Bild: <https://wwwpixabay.com>

Quelle Zitat: <https://www.cio.com/article/3003538/study-reveals-that-most-companies-are-failing-at-big-data.html>

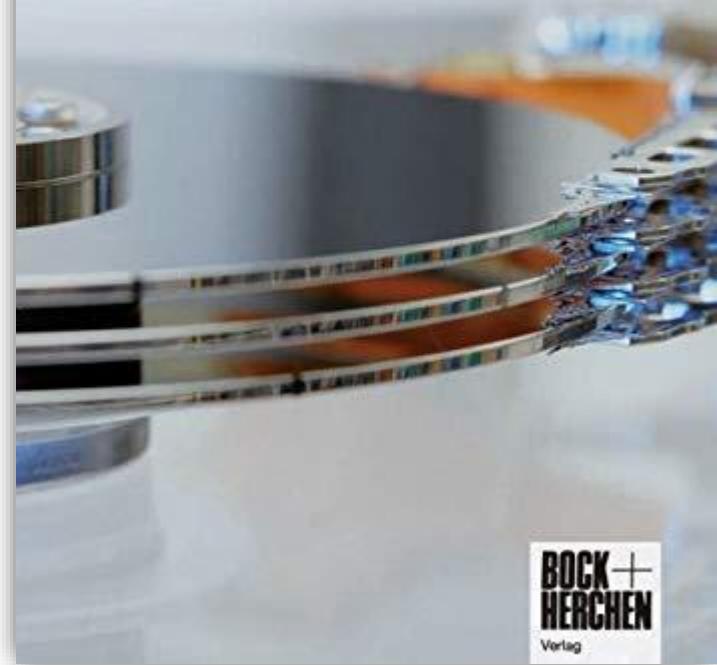
Inwertsetzung der Daten

„Riesige Mengen an Daten anzuhäufen macht nur dann Sinn, wenn es Hilfsmittel gibt, die in diesen Datenmengen enthaltenen **Schätze auch zu heben** – ansonsten entstehen lediglich **gigantische Datenfriedhöfe**.“

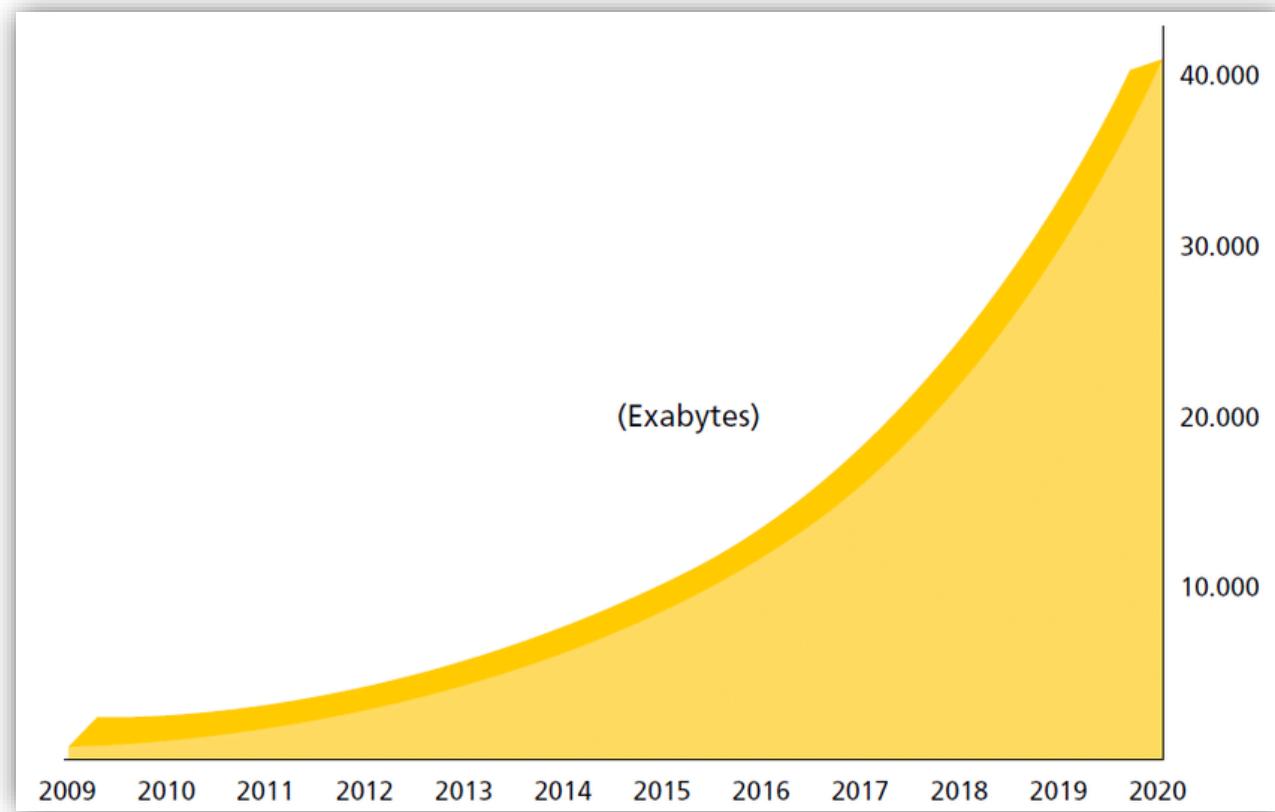
Handbuch Forschungsdatenmanagement
Büttner et al, 2011

Büttner / Hobohm / Müller (Hrsg.)

Handbuch Forschungsdaten- management



Menge an Daten



Quelle: https://www.researchgate.net/figure/The-exponential-data-growth-estimated-between-2010-and-2020-1_fig1_329298440

1 Gigabyte GB = 1000 MB
1 Terabyte TB = 1000 GB
1 Petabyte PB = 1000 TB
1 Exabyte EB = 1000 PB

Teil 1

Daten und
ihr Wert

Teil 2

Der Nutzen
eines DMP
für die
Forschungsarbeit



Datenmanagementplan (DMP)

DMP-Template EU Horizon 2020:

1. Data **summary**
2. **FAIR** Data
Findable
Accessible
Interoperable
Re-usable
3. Allocation of **resources**
4. Data **security**
5. **Ethical** aspects



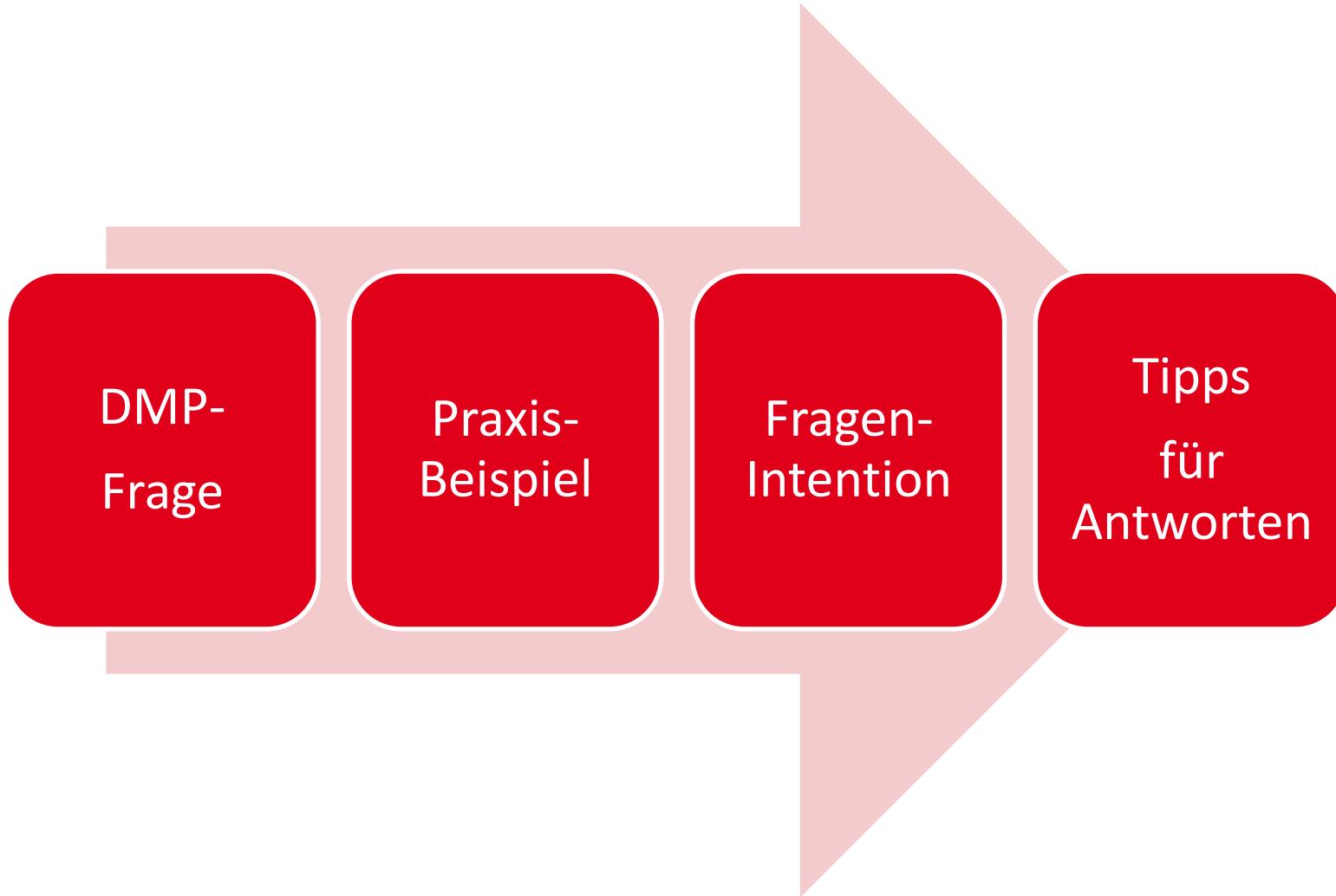
Quelle Text: [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/
open-access-data-management/data-management_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)
Quelle Bild: https://en.wikipedia.org/wiki/European_Commission

Datenmanagementplan (DMP)

DMP = *langweilig, glanzlos, dumpf,...*?

The screenshot shows a web page from the journal 'nature'. At the top, there is a red header bar with a 'MENU' dropdown, the 'nature' logo ('International journal of science'), and buttons for 'Subscribe', 'Search', and 'Login'. Below the header, the text 'EDITORIAL • 13 MARCH 2018' is visible. The main title of the article is 'Everyone needs a data-management plan'. Below the title, a subtitle reads 'They sound dull, but data-management plans are essential, and funders must explain why.' At the bottom of the article section, there is a link 'Quelle: <https://www.nature.com/articles/d41586-018-03065-z>'.

Datenmanagementplan (DMP)



1 Data summary

What is the expected size of the data?

1 Data summary

What is the expected size of the data?



Quelle: <https://www.nagra.ch/de/news/news-detail/bohrung-in-riniken-ermoeglicht-blick-in-die-juengere-geologische-vergangenheit-und-in-die-zukunft.htm>

1 Data summary

What is the expected size of the data?

Fragen-Intention:

- (insgesamt) zu erwartende Datenmenge
- Vermeidung von Datenverlust
- Bedarfsgerechte IT-Infrastruktur (inkl. Finanzbedarf)

1 Data summary

Will you re-use any existing data?

1 Data summary

Will you re-use any existing data?



New Findings from Old Data

Recalibrated and reanalyzed data from the Voyager flybys of Jupiter 40 years ago, presented in a series of papers in *JGR: Space Physics*, show the value of archival data.

Quelle: <https://eos.org/editors-vox/new-findings-from-old-data>

Mike Liemohn,
Editor of *JGR: Space Physics*, 2017

1 Data summary

Will you re-use any existing data?

Fragen-Intention:

- (Kosten-) Effizienz der Forschung
- Effektivität der Forschung
- Kenntnis über Forschungslandschaft (wer verfügt über welche Daten)

2 FAIR data - 2.1 Making data findable

What naming conventions do you follow?

2 FAIR data - 2.1 Making data findable

What naming conventions do you follow?

Doing the right thing: Authors retract brain paper with “systematic human error in coding”

A group of Swiss neurologists have lost their 2013 article in Frontiers in Human Neuroscience after reporting that their data were rendered null by coding errors.



Quelle: <https://retractionwatch.com/2014/01/07/doing-the-right-thing-authors-retract-brain-paper-with-systematic-human-error-in-coding/>

2 FAIR data - 2.1 Making data findable

What naming conventions do you follow?

authors discovered that a systematic human error in coding the name of the files had been made during the extraction of the EEG template topographic maps best differentiating the two

...

error has ultimately modified the final result of the paper, which is therefore not correct. For this reason, the authors request to withdraw the article.

Quelle: <https://retractionwatch.com/2014/01/07/doing-the-right-thing-authors-retract-brain-paper-with-systematic-human-error-in-coding/>

2 FAIR data - 2.1 Making data findable

What naming conventions do you follow?

Fragen-Intention:

- Einheitlichkeit & konsequente Anwendung
- (idealerweise) sprechende Dateinamen
- Beispiel:
 - *2019-09-19_vonderDunk_DMP-Präsentation_SaxFDM-Tagung.pptx*
 - [YYYY-MM-DD_Autor_Thema]

2 FAIR data - 2.1 Making data findable

What metadata will be created?

2 FAIR data - 2.1 Making data findable

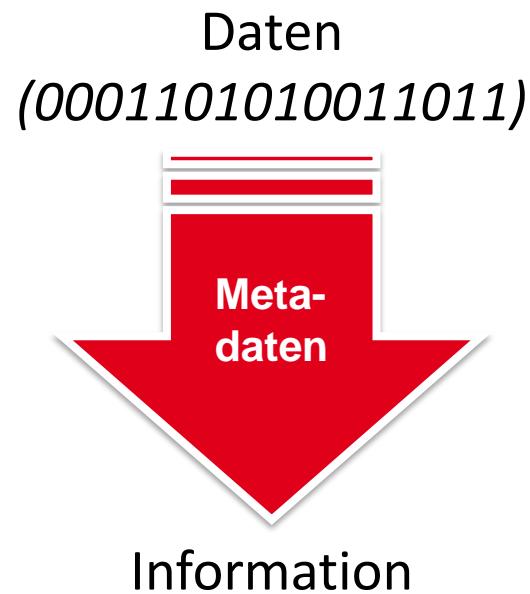
What metadata will be created?



Quelle: <https://libraries.mit.edu/data-management/services/workshops/>

2 FAIR data - 2.1 Making data findable

What metadata will be created?



2 FAIR data - 2.1 Making data findable

What metadata will be created?

Fragen-Intention:

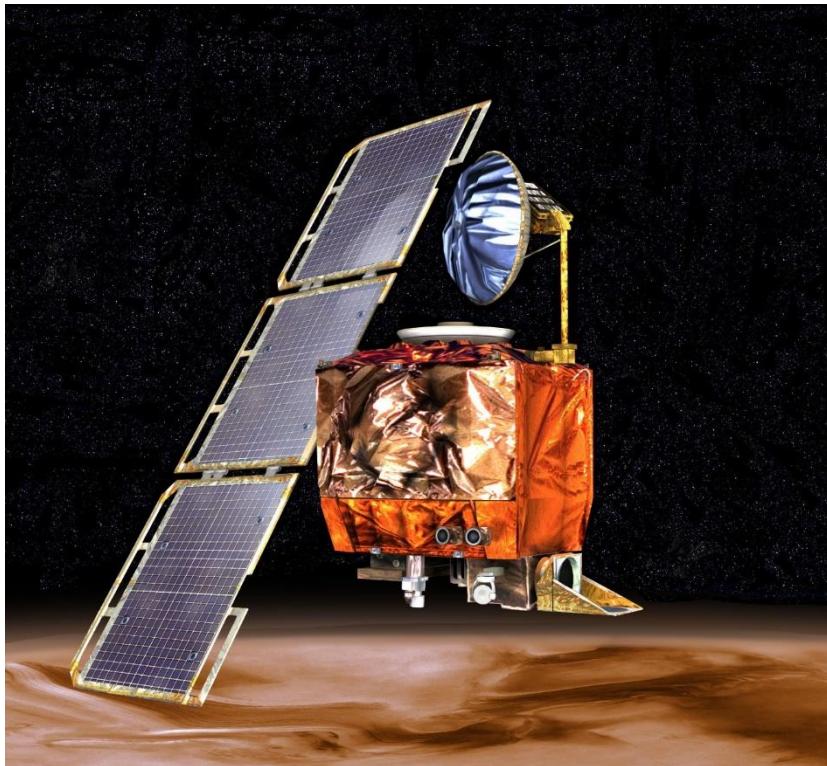
- Metadaten begleiten das Projekt von Anfang bis Ende
- Verwendung von Metadaten-Standards (*z.B. Dublin Core, DataCite*)
- Erkenntnis: der dankbarste Nutzer der Metadaten ist man oft selber

2 FAIR data - 2.3 Making data interoperable

Are the data produced interoperable?

2 FAIR data - 2.3 Making data interoperable

Are the data produced interoperable?



Quelle: https://en.wikipedia.org/wiki/Mars_Climate_Orbiter

2 FAIR data - 2.3 Making data interoperable

Are the data produced interoperable?

Fragen-Intention:

- Verwendung interoperabler Dateiformate und Maßeinheiten
- Erstellung eines *Units & Conversions Fact Sheet*
- Zugänglichmachung des *Units & Conversions Fact Sheet* an zentraler Stelle

2 FAIR data - 2.4 Increase data re-use

How long does the data remain re-usable?

2 FAIR data - 2.4 Increase data re-use

How long does the data remain re-usable?



Quelle: <https://www.tapestodigital.com/>

2 FAIR data - 2.4 Increase data re-use

How long does the data remain re-usable?

Fragen-Intention:

- Erkenntnis: Speichermedien veralten
- Einschätzung bzgl. Lebensdauer der verwendeten Speichermedien
- Ggf. Maßnahmen zur Überspielung auf aktuelle Speichermedien (inkl. Kostenschätzung)

2 FAIR data - 2.4 Increase data re-use

Are data quality assurance processes described?

2 FAIR data - 2.4 Increase data re-use

Are data quality assurance processes described?



Quelle: <https://www.bbc.com/news/world-europe-48134881>

2 FAIR data - 2.4 Increase data re-use

Are data quality assurance processes described?

Fragen-Intention:

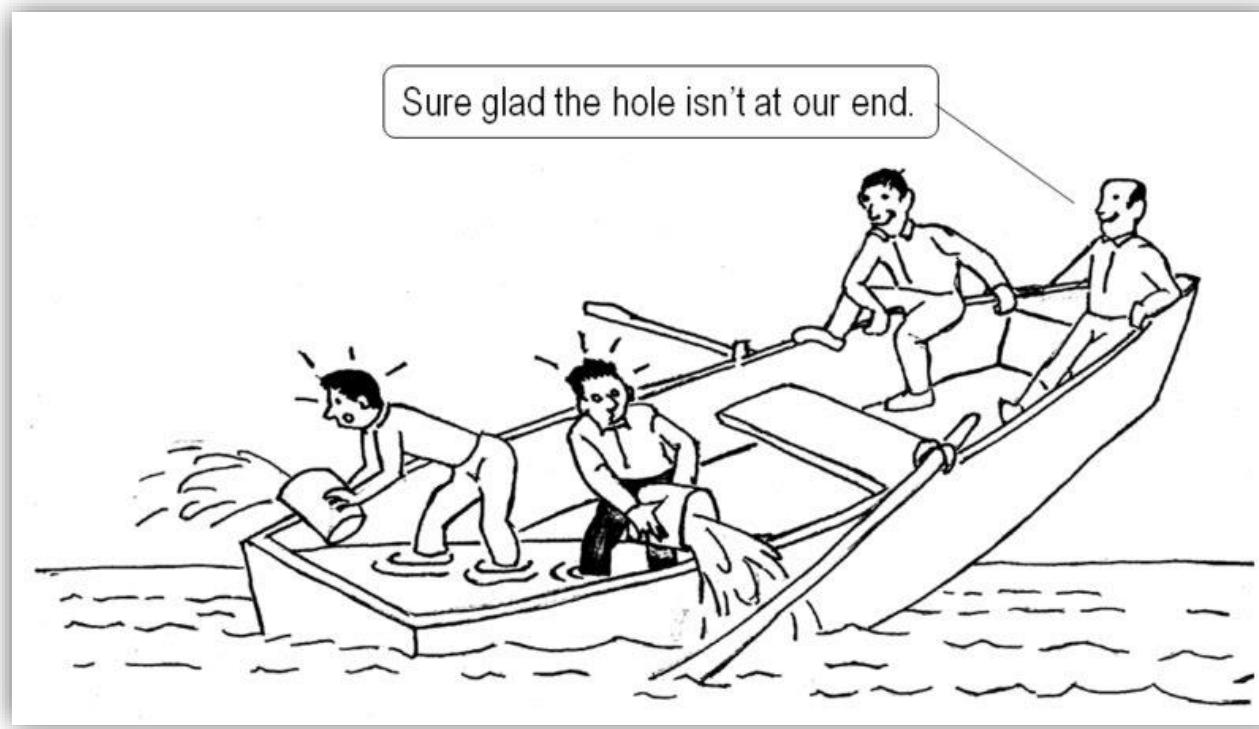
- Interner Daten-Review (*4-Augen-Prinzip*)
- (idealerweise) Externer Review
- Qualitätsrichtlinien (*z.B. Validierung der Datenaktualität*)

3 Allocation of resources

Who will be responsible for data management?

3 Allocation of resources

Who will be responsible for data management?



Quelle: <https://yhsjo.files.wordpress.com/2012/11/responsibility2.jpg>

3 Allocation of resources

Who will be responsible for data management?

Fragen-Intention:

- FDM = Gruppenaktivität
- Innerhalb der Gruppe: klare Zuweisung von Verantwortlichkeiten, z.B.
 - Forschende → inhaltliche Korrektheit, Metadaten
 - Data Librarian → logische Kontrolle der Datenpakete
 - Rechenzentrum → Bereitstellung der IT-Infrastruktur

4 Data Security

What provisions are in place for data security?

4 Data Security

What provisions are in place for data security?

13. August 2018, 13:41 Uhr Nordrhein-Westfalen

JVA-Beamter verliert USB-Stick mit sensiblen Daten

- Ein Mitarbeiter der Justizvollzugsanstalt Euskirchen hat einen Stick mit sensiblen Daten auf dem dortigen Parkplatz verloren.
- Ein Häftling soll ihn eingesteckt haben, bis jetzt haben die Beamten den Stick aber noch nicht gefunden.

Quelle: <https://www.sueddeutsche.de/panorama/nordrhein-westfalen-jva-beamter-verliert-usb-stick-mit-sensiblen-daten-1.4091471>

4 Data Security

What provisions are in place for data security?

Fragen-Intention:

- Back-Up (USB-Sticks sind kein Back-Up)
- 3-2-1-Backup-Regel
 - 3 Datenkopien (1x Kopie der Rohdaten, 2x Back-Up der Kopie)
 - 2 Medien (Kopien auf min. 2 unterschiedlichen Speichermedien ablegen)
 - 1 externes Backup (Original und Back-Ups nicht an 1 physischen Ort aufbewahren)



Quelle: <https://pixabay.com>

5 Ethical aspects

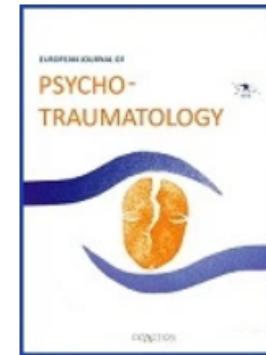
Are ethical and legal issues being considered?

5 Ethical aspects

Are ethical and legal issues being considered?

Privacy breach prompts retraction of three papers from the trauma literature

A group of international psychology researchers is retracting three papers in the wake of revelations that they failed to adequately safeguard the identities of the patients who participated in the studies.



Quelle: <https://retractionwatch.com/2013/07/31/privacy-breach-prompts-retraction-of-three-papers-from-the-trauma-literature/>

5 Ethical aspects

Are ethical and legal issues being considered?

oversigh[t]s. However, more importantly, the auditors found out that data files contained names and dates of birth of all participants. Moreover, these data files had been sent between study sites, i.e., Zurich, Berlin, and Southampton. In Switzerland (and I assume in other countries as well), this clearly constitutes a breach of the data privacy act.

Quelle: <https://retractionwatch.com/2013/07/31/privacy-breach-prompts-retraction-of-three-papers-from-the-trauma-literature>

5 Ethical aspects

Are ethical and legal issues being considered?

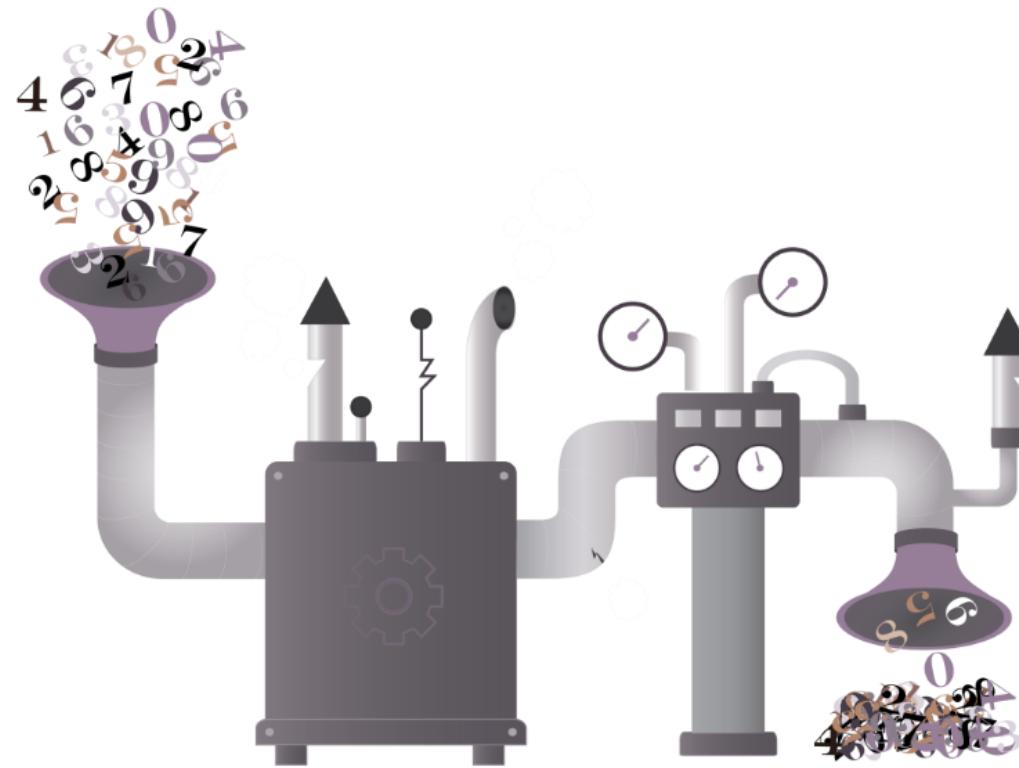
Fragen-Intention:

- Ausreichende Anonymisierung personenbezogener Daten
- Beachtung der DSGVO
- Beachtung von *Intellectual Property Rights*

Fazit

Fazit

Garbage in – Garbage Out



Quelle: <https://www.cleanpng.com/png-garbage-in-garbage-out-management-computer-data-co-1816657/download-png.html>