

# gesis

Leibniz-Institut  
für Sozialwissenschaften



## Einsatz von Software im Forschungsdatenmanagement

Reiner Mauer, GESIS

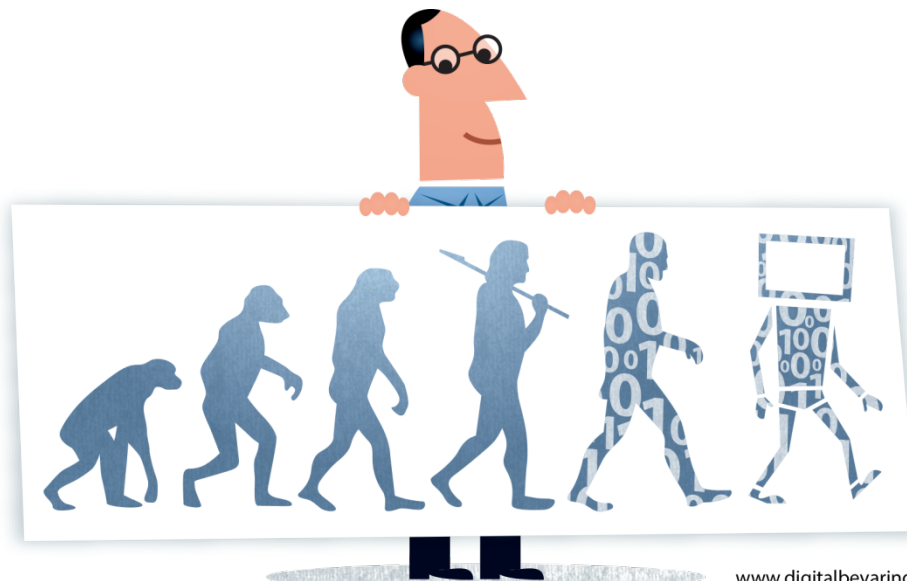
*7. Workshop des AK Forschungsdaten der Leibniz-  
Gemeinschaft, Berlin, 11.06.2018*

# Inhalt

1. Einsatz von Software im Forschungsdatenmanagement
2. Auswahl und Nutzung
  - i. Rahmenbedingungen und Anforderungen
  - ii. Planung
  - iii. Formate
  - iv. Eigenentwicklung
  - v. Open Source
3. Fazit/ Zum Mitnehmen

# (1) Einsatz von Software im FDM

- Zunehmender Einsatz von Software in allen Phasen des Forschungsprozesses zur Generierung/Erhebung/Messung, Bearbeitung, Analyse und Visualisierung von Forschungsdaten
- Technologische Entwicklung dynamisch bis rasant



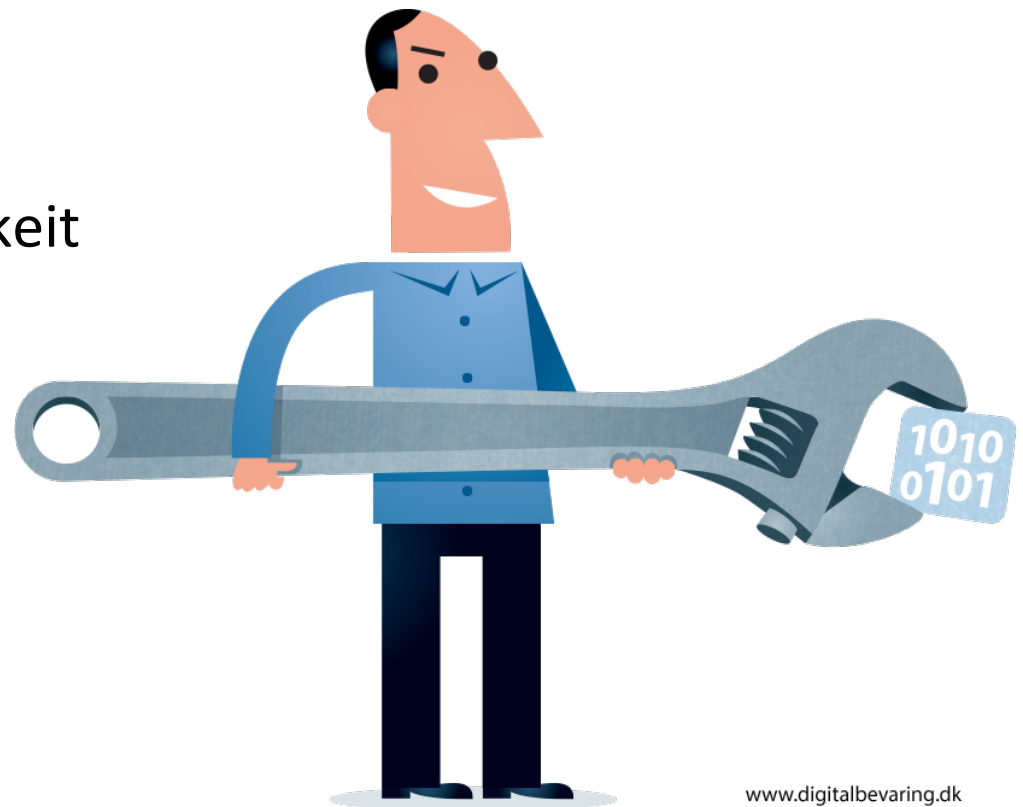
# (1) Einsatz von Software im FDM

- gleichzeitig steigen Anforderungen an Umgang mit Forschungsdaten, bspw. in Bezug auf
  - ▶ Zugänglichkeit
  - ▶ Nachnutzbarkeit
  - ▶ Transparenz, Nachvollziehbarkeit u. Reproduzierbarkeit von Forschungsprozessen u. –ergebnissen



# (1) Einsatz von Software im FDM

- enger Zusammenhang zwischen Forschungsdaten und Software
  - ▶ (Nach)Nutzbarkeit
  - ▶ Kollaborationsfähigkeit
  - ▶ Anschlussfähigkeit
  - ▶ Reproduzierbarkeit
  - ▶ Rechtskonformität
  - ▶ ...



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

# (1) Software im FDM-Prozess

- „Schweizer Messer“ für Forschungsdaten gibt es leider nicht
  - ▶ Software ist i.d.R. auf bestimmte Datentypen, Methoden, Phasen, Prozessabläufe, Anforderungen oder Aufgaben zugeschnitten
  - ▶ daher auch keine Vorstellung von konkreten Softwareprodukten im Sinne von Empfehlungen
  - ▶ Software, die unmittelbar zur Erzeugung/ Erfassung/ Messung und Verarbeitung von Forschungsdaten eingesetzt wird
  - ▶ Software, die (allgemeine) Datenmanagementaufgaben unterstützt, bspw. Dokumentation, Versionierung, Speicherung, Datenaustausch oder Planung des FDM selbst
- Projekte/ Forschungsgruppen/ Institute müssen **eigenes Portfolio definieren**
- Auswahl u. Einsatz von Software erfordert **gezielte Planung und konstantes Monitoring** (Anforderungen, Ziele, Workflows, Umfeld)



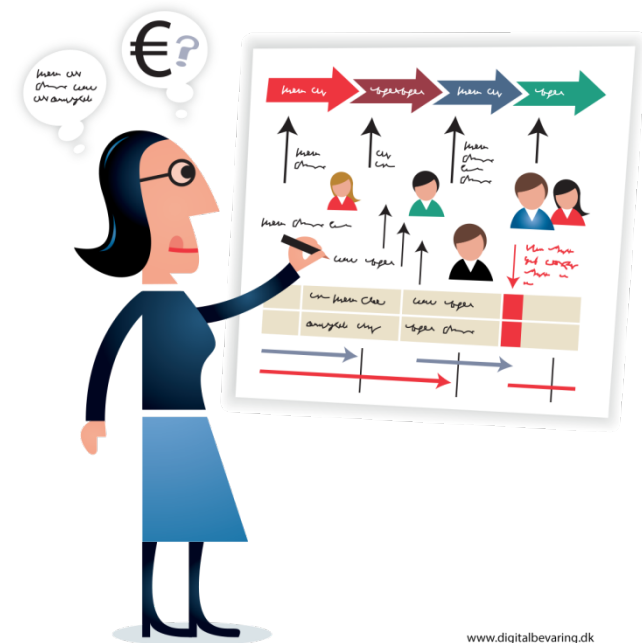
## (2) Auswahl u. Nutzung: Planung

- Einstieg sollte aus Planung des Workflows bestehen, u.a.

- ▶ Aufgaben und Arbeitsschritte
- ▶ Beteiligte (viele? international?)
  - Datenaustausch, Zugriffskonzepte, Versionierung, ....
  - Projekt-/Wissensorganisation
- ▶ Datenformate (insbesondere mit Blick auf (Nach)Nutzungsszenarien)
- ▶ Relevante Standards
- ▶ Nutzungsszenarien

→ Datenmanagementplan bietet sehr gute Grundlage für diese Planung

→ Gibt es etablierte Tools, Workflows, Standards, Best Practices in der Community?



## (2) Auswahl u. Nutzung: Anforderungen

**Anforderungen von Förderern, Politik, Journalen sowie gesetzliche Auflagen spielen wichtige Rolle, u.a.:**

- Nachvollziehbarkeit und Transparenz von Forschungsmethoden, Arbeitsprozessen und –ergebnissen als essentieller Bestandteil wissenschaftlicher Integrität (WR 2015)
  - ▶ Dokumentation, Aufbewahrung, Nutzbarkeit und Zugänglichkeit
- Forschungsdaten sollen „FAIR“ sein
  - ▶ auffindbare, zugängliche, interoperable und nachnutzbare (Meta)Daten
- Verarbeitung personenbezogener Daten
- Urheberrechte, Lizenzen
- Dokumentations- und Aufbewahrungspflichten



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

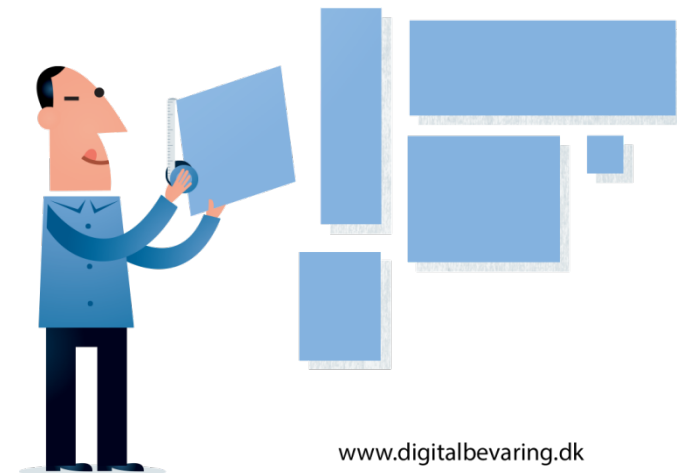


## (2) Auswahl u. Nutzung: Anforderungen

- Sicherung von Datenqualität
  - ▶ Dokumentation spielt zentrale Rolle bei **Bewertung und Nachweis** von Qualität
  - ▶ idealerweise automatisiert zum Zeitpunkt der Messung bzw. Bearbeitung und in standardisierter Form (Metadaten); spätere Erfassung aufwändiger u. teurer
  - ▶ Z.B.: Wer hat wann, zu welchem Zweck, was und womit an den Daten verändert
  - ▶ Daten selbst sollten natürlich auch einer Qualitätskontrolle unterzogen werden , z.B. durch automatisierte Plausibilisierung von Messwerten
  - ▶ Beispiele für weitere Maßnahmen: Versionierung, Checksummen
- Kollaboration und Austausch ermöglichen
- Risiken beim Umgang mit Daten minimieren (Datenverlust oder –korruption, unberechtigte Zugriffe, etc.)
- ➔ Bei der Planung nicht nur die unmittelbaren Projektziele, sondern auch übergeordnete Rahmenbedingungen und Anforderungen beachten

## (2) Auswahl u. Nutzung: Formate

- Wichtiges Kriterium bei der Auswahl von Software sind die damit verbundenen Dateiformate:
  - unbedingt ganzen Forschungsprozess (inklusive Nachnutzung) im Blick haben
  - Sind die Daten unabhängig von Software nutzbar? Lassen sie sich leicht austauschen und von anderen nutzen?
  - Werden Formate bei Softwareaktualisierungen ebenfalls verändert? Können sie dann noch von älteren Versionen gelesen werden?
  - Können (proprietäre) Formate verlustfrei in andere (offene) Formate überführt werden?
- Offene und weitverbreitete Formate sind proprietären vorzuziehen
- Orientierung an Community-Formaten
- Sollten Daten nicht verlustfrei migrierbar sein, muss Software erhalten werden



### (3) Auswahl u. Nutzung: Eigenentwicklung

- Eigenentwicklung bietet Möglichkeit Software komplett auf spezifische Bedarfe auszurichten, aber
  - ▶ Veränderte Anforderungen u. hohe Dynamik im Umfeld (IT, Standards, Recht, etc.) erfordern **kontinuierliche Pflege, Anpassung und Weiterentwicklung** und binden entsprechende Ressourcen
  - ▶ Entsprechende Expertise muss vorhanden sein und gehalten werden (wissenschaftl. Umfeld geprägt von befristeten Arbeitsverträgen, Markt f. Software-Entwickler\_innen ist hochkompetitiv)
  - ▶ Gefahr von „Insellösungen“



*“If you want  
a thing done  
well, do it  
yourself.”*

## (2) Auswahl u. Nutzung: Open Source

- frei verfügbare, quelloffene Software bietet viele Vorteile; aber auch hier Bewertung notwendig
- mögliche Kriterien zur Beurteilung:
  - ▶ Reputation
  - ▶ Entwicklungsaktivität
  - ▶ Verwendung von Standards u. Interoperabilität
  - ▶ Support (Community u./od. kommerziell)
  - ▶ Updates/ Bugfixing
  - ▶ Dokumentation
  - ▶ Kompetenzen im eigenen Haus vorhanden
  - ▶ geordneter Entwicklungsprozess erkennbar
  - ▶ rechtliche Rahmenbedingungen der Nutzung (Lizenz)

„Top tips for selecting open source software“, OSSWatch; University of Oxford

# Zum Mitnehmen



- Einsatz von Software sorgfältig planen
- Erfahrungen von anderen Anwender\_innen einholen
  - ▶ Limitationen, Probleme usw. finden sich nicht unbedingt in Handbüchern
- Systemabhängigkeit verhindern!
  - ▶ Forschungsdaten sowie alle verbundenen Objekte und Informationen sollten unabhängig von aktuell eingesetzter Software nutzbar sein
- Einsatz von Open Source Software ist grundsätzlich sehr gute Strategie, die im konkreten Einzelfall aber auch immer überprüft werden sollte
- Schulungs-, Beratungs- u. Unterstützungsbedarfe von Forschenden hinsichtlich Auswahl und Nutzung von Software adressieren

Kontakt: [reiner.mauer@gesis.org](mailto:reiner.mauer@gesis.org)

Vielen Dank für Ihre Aufmerksamkeit!

gesis

Leibniz-Institut  
für Sozialwissenschaften

Mitglied der  
*Leibniz*  
Leibniz-Gemeinschaft

## Quellen

- Wissenschaftsrat (2015): Empfehlungen zu wissenschaftlicher Integrität.  
<https://www.wissenschaftsrat.de/download/archiv/4609-15.pdf>
- OSS-Watch: Top tips for selecting open source software;  
<http://oss-watch.ac.uk/resources/tips> (zuletzt aufgerufen 10.06.2018)

## Bilder und Grafiken

- Bild Swiss Army Knife (CC0). Quelle:  
<https://pixabay.com/de/schweizer-taschenmesser-werkzeuge-572667/>
- Napoleon (Gemeinfrei), <https://commons.wikimedia.org/w/index.php?curid=22174172>
- Alle anderen Illustrationen: [www.digitalbevaring.dk](http://www.digitalbevaring.dk) (CC BY 2.5. Dänemark Lizenz );  
Wenn nicht anders angegeben, ist Jørgen Stamp der Autor der Illustrationen.