

Sicherstellung der Reproduzierbarkeit von  
Forschungsergebnissen durch Bewahrung  
des Zugriffs auf Forschungssoftware  
**Forschungssoftware Managen**

Dirk von Suchodoletz

15.09.2022 – DINI Workshop in Stuttgart

# Problemaufriss

- BioDATEN Teil der Science Data Center in BaWü
- Ziel der Unterstützung der Bio-Informatik-Community, Klärung rechtlicher Aspekte
- FAIR impliziert Mitbetrachtung von Software, die Bestandteil von Forschungsprozessen
- Bestandteil des Kontextes von Forschungsdaten/-prozessen

# Wiederherstellung von Kontexten

- Glaubwürdigkeit der Wissenschaft erfordert gute Reproduzierbarkeit
- Reproduzierbarkeit von Ergebnissen und Workflows nicht nur im Moment der Erstellung, Verarbeitung von Daten
- Herausforderung bereits für Review-Prozesse durch Dritte, die auch entsprechende Software benötigen
- Besondere Herausforderung: Langfristige Reproduzier- und Nachnutzbarkeit

# Versuch einer Klassifikation

- Mögliche Betrachtungsmetriken
  - Quelloffene (insbes. Open Source) vs. Herstellergebunden
  - Forschungssoftware vs. Software für Forschende
  - SW-Framework, App, Skript, Bibliothek
- Je nach Software spezielle Herausforderungen
  - Viele Überlegungen bereits zu Open Source und technischen Systemen (Vortrag eines Kollegen)
  - Hier primär Betrachtung der rechtlichen Rahmenbedingungen und organisatorische Aspekte

# Herstellergebundene SW

- Herausforderung Lizenzen
  - Bedingungen zunehmend komplexer
  - Spezielle Belange von Forschung typischerweise nicht berücksichtigt
  - Deutliche Verkürzung der Laufzeiten, Abo-Modelle
- Weitergabe an Dritte schwierig
- Mitarchivierung z.B. auf Datenpublikationsrepository kaum möglich
- Lizenzen und Versionen im Zeitablauf (mit oft starken Änderungen)
- Nichtvorhandene Lizenzen können Nachnutzung, Reproduktion verhindern (insbesondere durch wirtschaftlich schwächere Teilnehmende im Prozess)

# Herstellergebundene SW

- Schwer möglich: Aufbau öffentlicher Reproducibility services
- Abhängigkeit von Produktpolitik des Herstellers
- Zusätzlich: “Enigma-Charakter” der Software
- Weitere Risiken durch nicht dokumentierte Änderungen über Versionen bzw. Updates hinweg
- Herstellergebundene Dateiformate, die adäquate Betrachtung durch Drittsoftware, spätere Nachnutzung effektiv verhindern können

# Herstellergebundene SW

- Insgesamt
  - Zunehmend problematisches Vorgehen der Hersteller
  - Verstärkte Marktkonzentration reduziert Verhandlungsmacht der Wissenschaftsseite
- Neue technische Entwicklungen verlagert Installation und Ausführung von Software zunehmend in die Cloud (zum Anbieter)

# Empfehlungen

- Stärkerer Fokus auf quelloffene Software
  - Bevorzugung in Förderbedingungen (der Forschungsförderer)
  - Verankerung über institutionelle Policies (Universität, Forschungseinrichtung, Projekt)
  - Best-Practices für Software angelehnt an FAIR-Prinzipien
  - Förderung des Einsatzes in der Lehre
- Schon beim Setup von Projekten durch geeignete Auswahl darauf achten (z.B. via DMP Verwendung von Software transparent machen)
- Bei Bereitstellung eigens entwickelter/eigener Software – Prüfung auf generelle Freigabe/Veröffentlichung

# Empfehlungen

- Bei unvermeidbarer herstellergebundener Software
  - Zentral: Starkes gemeinsames Auftreten
  - Förderung von Alternativen (z.B. MatLab → Python)
  - Forderung nach entsprechenden Lizenzbedingungen für Reproduzierbarkeit
  - Berücksichtigung technisch-ethischer Fragen wie z.B. Überwachbarkeit von Nutzenden
  - Forderung beispielsweise auch nach Offenlegung nach gewisser Zeit
- Verbesserung von Transparenz, Erlauben von Code-Reviews

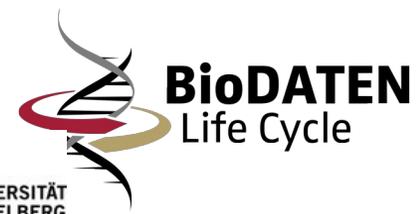
# Empfehlungen

- Hinterlegung von Software an zentraler Stelle
  - Escrow Services
  - “Legal deposit” beispielsweise bei Nationalbibliothek oder passenden Spezial-/Forschungsbibliotheken
  - Abstimmung des Betriebs gemeinsamer Repositories
- NFDI schafft hier neue Möglichkeiten/Chancen
  - Plattform für gemeinsames, abgestimmtes Vorgehen
  - Sektion Common Infrastructures / Base4NFDI
  - Rechtliche Beratung und Unterstützung (Sektion ELSA)

# Ausblick

- Internationale Entwicklungen bieten Anschauung für weiteres Vorgehen
  - Zentrale/gemeinsame Repositories für relevante Software
  - Vereinbarungen mit den SW-Herstellern für Nutzung für Reproduktion von Forschung
  - Kooperation mit internationalen Initiativen sinnvoll

# Thank you for your attention!



UNIVERSITÄT  
HOHENHEIM



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



BioDATEN is  
sponsored by



Baden-Württemberg



MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST