

Research Data Publication using InvenioRDM

Jonathan Bauer

BioDATEN Abschluss-Workshop in Tübingen

12. Juni 2023

Anforderungen an ein Repository

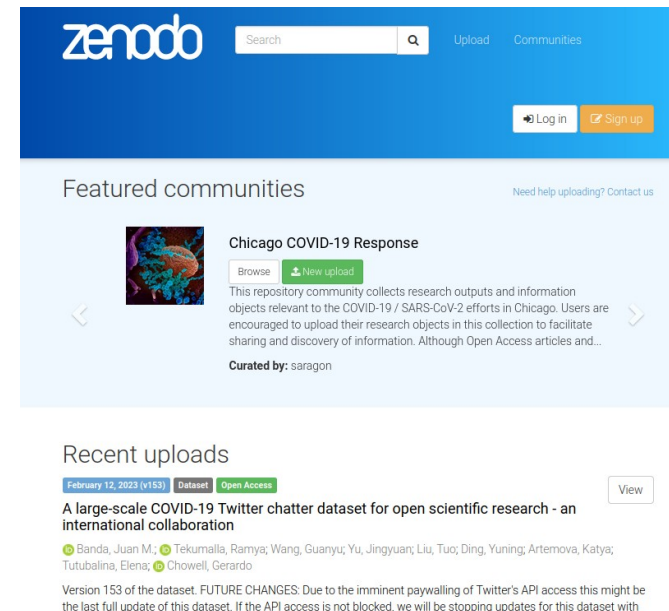
Publikation von Forschungsdaten erlangt steigende Bedeutung für Einrichtung und ihre Forschenden -> *Bedarf nach geeigneter Plattform*

- Umgang mit großen Datenpaketen, die nicht in klassischen UB-Publikationssystemen eingestellt werden können
 - Abbildung community-spezifischer Metadaten und Veröffentlichungs-Workflows
 - Anpassungsfähigkeit für verschiedene Projekte und Aufgaben
 - Zitierfähige Veröffentlichung von Forschungsdaten
 - Nachhaltige Perspektive und Entwicklung
 - Anbindung eines zukunftsfähigen und sicheren Speicher-Backends
 - Automatisierte Bereitstellung von Metadaten und Publikationsinformationen
- > Erarbeitung eines [Kriterienkatalogs](#) mit BW SDCs und im Rahmen des AK FDM

InvenioRDM – Fact Sheet

Auswahl von [InvenioRDM](#):

- [Zenodo.org](#) als Framework-Implementierung
- Nachhaltige Entwicklerperspektive (CERN)
 - Open Source Framework / große Community
 - Sehr aktive Entwicklung mit Eingehen auf Nutzerwünsche
- Internationale Community
 - TÜ und FR offizielle Partner und Teil der Community Meetings
- Aufbau institutioneller Repositorien an Uni Freiburg und Uni Tübingen
- Ermöglicht mehrere fachspezifische Communities in einer Instanz
- Einsatz bei SDC BioDATEN und NFDI DataPLANT



InvenioRDM – Schnittstellen

Auswahl von [InvenioRDM](#):

- Anbindung an DataCite -> DOI-Registrierung
- Anbindung an ORCID -> Übernahme von Datenpublikationen
- Anbindung an z.B. FIS oder andere Systeme via REST-API
- Bereitstellung von Object Storage über bwSFS
- Bereitstellung von Publikationsinformationen und Metadaten über OAI-PMH
- Anbindung mehrerer AAI / IDP via OpenID-Connect & SAML

Rechtlicher Rahmen

Nutzung der ORCID

- Abstimmung in [BW zur Nutzung der ORCID](#)

Datenüberlassungsverträge

- Ausarbeitung und Integration von Datenüberlassungsverträgen in Publikationsworkflow

Auswahl geeigneter Lizenzen für Veröffentlichung

- Ziel: Unterstützen der Open Science Policy der Einrichtungen
- [Empfehlung CC-BY für Daten und CC0](#) für Metadaten (AK FDM)

Herausforderungen

Anbindung an DataCite

- Erhebung notwendiger [Metadaten](#) für DOI-Registrierung
- Absprache mit UB hinsichtlich Menge und Kostengerüst für DOI-Registrierung
- Limitierung / Steuerung von Datensätzen für DOI-Registrierung
- Qualitätssicherung der Veröffentlichung
- Sicherstellung / Zusicherung der langfristigen Verfügbarkeit der DOIs

Nutzung bwSFS (förderiertes Speichersystem für die Forschung in BaWü)

- Abstimmung mit den beteiligten Rechenzentren

Langfristiger Betrieb und Wartung des Repositoriums

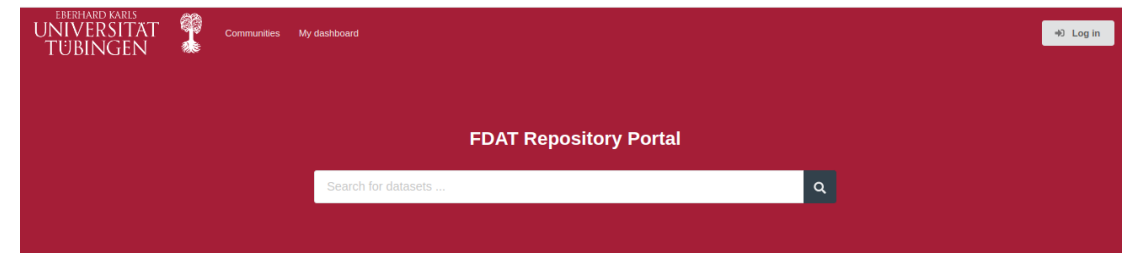
Austausch und Zusammenarbeit

- Gemeinsamer Rahmen zum Austausch zu Entwicklung und Administration seit mehreren Jahren
 - Gegenseitige Unterstützung bei Tests und neuen Anforderungen, wie beispielsweise das Community-Feature
 - Deployment-Workflows
 - Konfigurationsüberlegungen
- Enge Abstimmung mit den Bibliotheken
- Koordination in der Nutzung eines föderierten Speichersystems
- Expertise aus BioDATEN floss in anderen Projekten (DataPLANT)

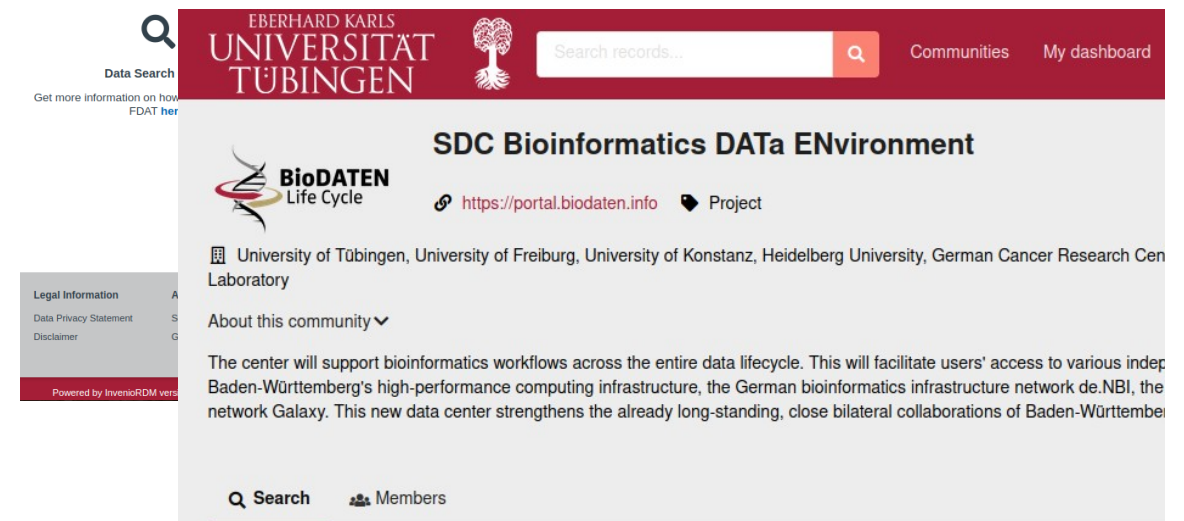
InvenioRDM als institutionelles Repository

Einsatz von InvenioRDM an der Universität Tübingen

- Generische Infrastruktur, die den Anforderungen von Drittmittelgebern entspricht
- Institutionelles Repository als disziplinenübergreifendes Angebot zum FDM
- Betrieb am Digital Humanities Center im mit Beratungsschwerpunkt auf Geistes- und Sozialwissenschaften, um nachhaltigen Umgang mit Forschungsdaten zu gewähren
- **Dedizierte BioDATEN Community**



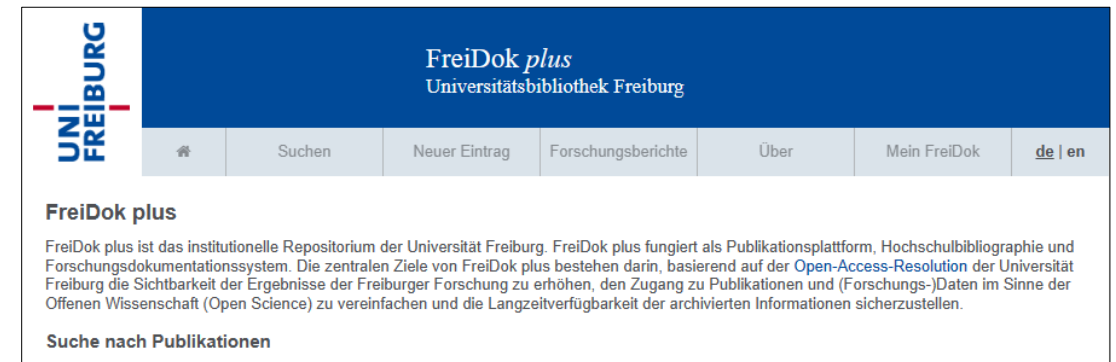
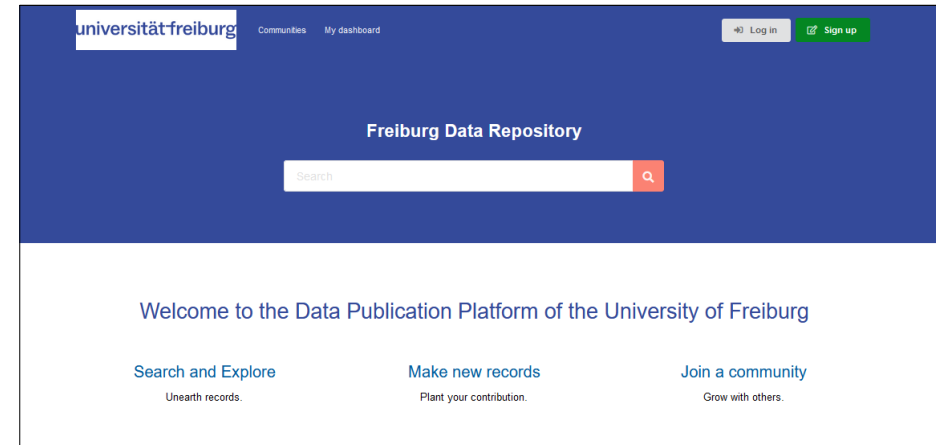
Welcome to FDAT - The Institutional Research Data Repository



InvenioRDM als institutionelles Repository

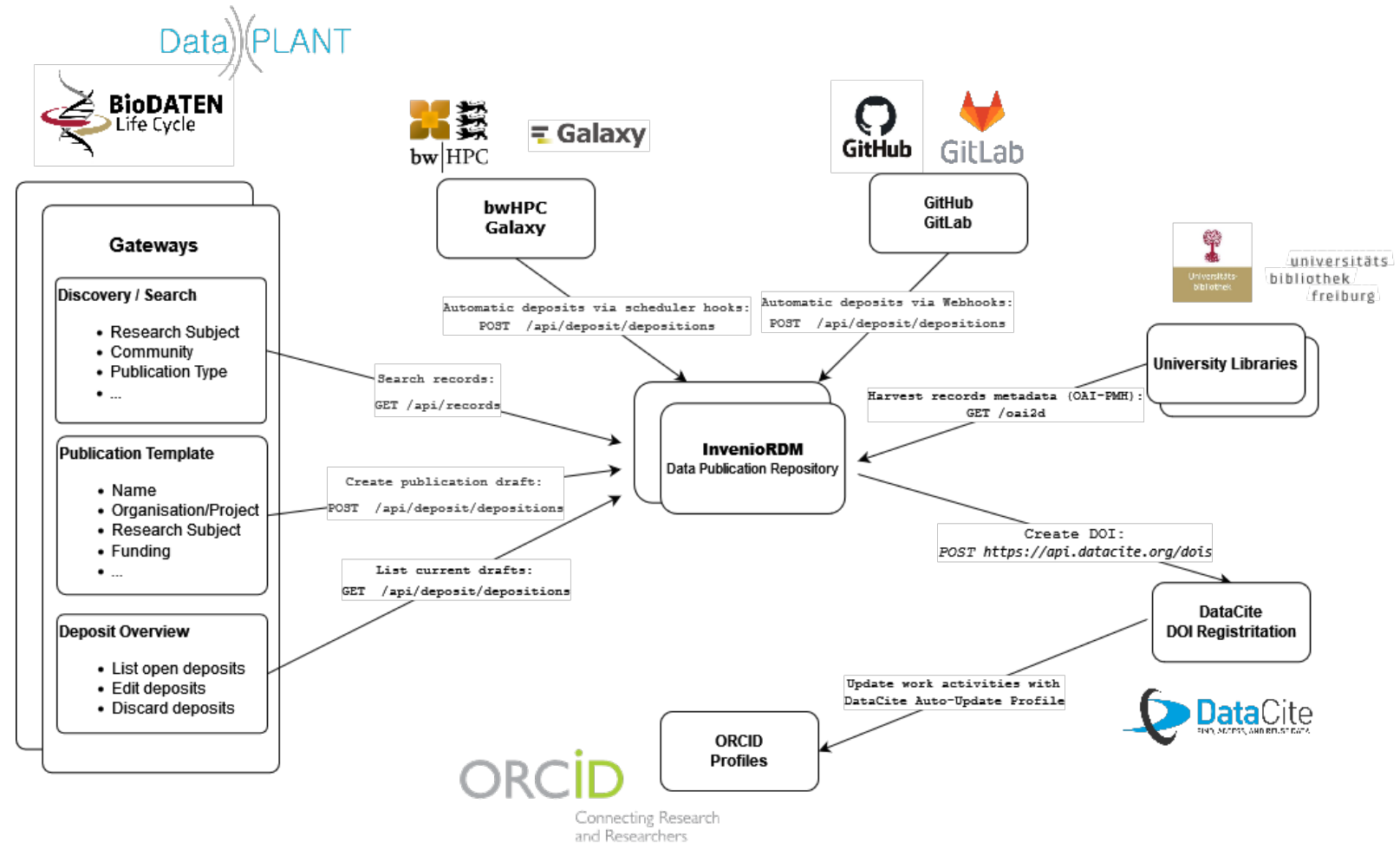
Einsatz von InvenioRDM an der Universität Freiburg

- Service der RDMG
 - Bedarfe verschiedener Exzellenz-Cluster und Projekte
 - Allgemeine, disziplinübergreifende Bedarfe der Forschenden ohne bestehende Community-Repositoryn
- Ergänzung des bestehenden FreiDok+ um Repository für Forschungsdaten



Integration von InvenioRDM

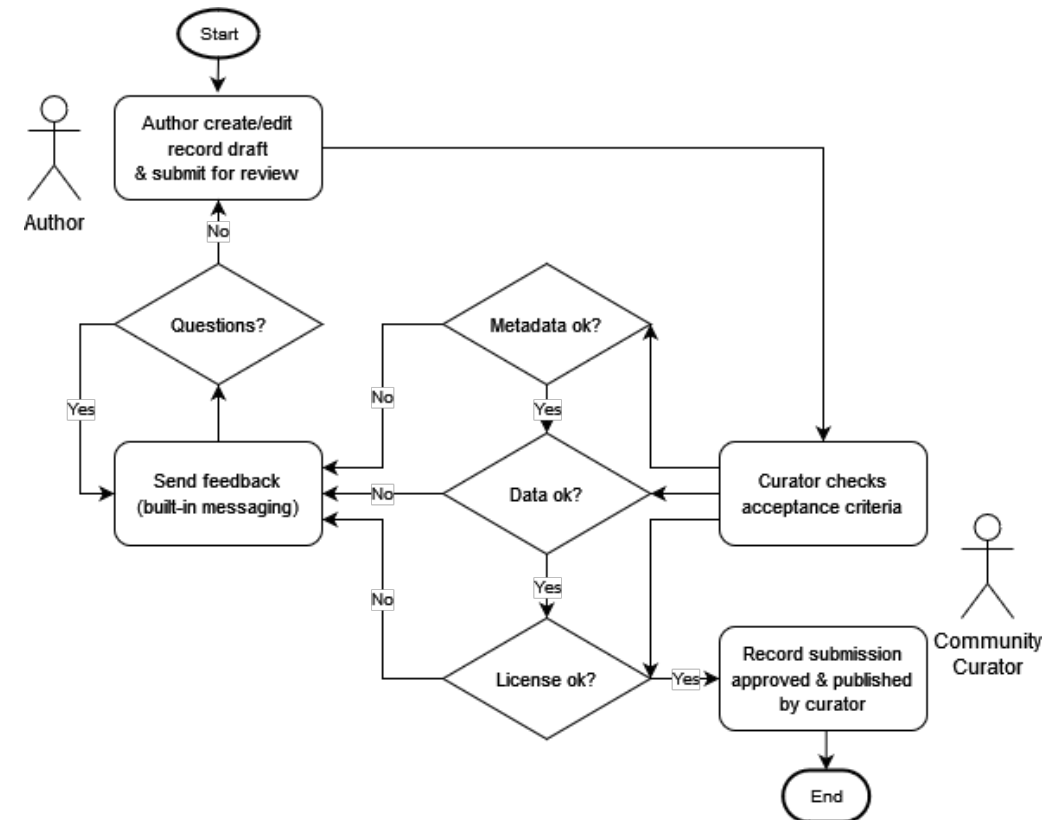
- Vielseitige Integrationen in Drittsystemen durch die REST-API
- Datenhaltungsplattformen
 - HPC-Cluster NEMO/BinAC
 - Galaxy
 - GitLab
- Forschungsinformationssysteme der UBs
- Durch DataCite zitierbare DOIs und automatisches Ergänzen des ORCID-Profiles



Workflows mit InvenioRDM

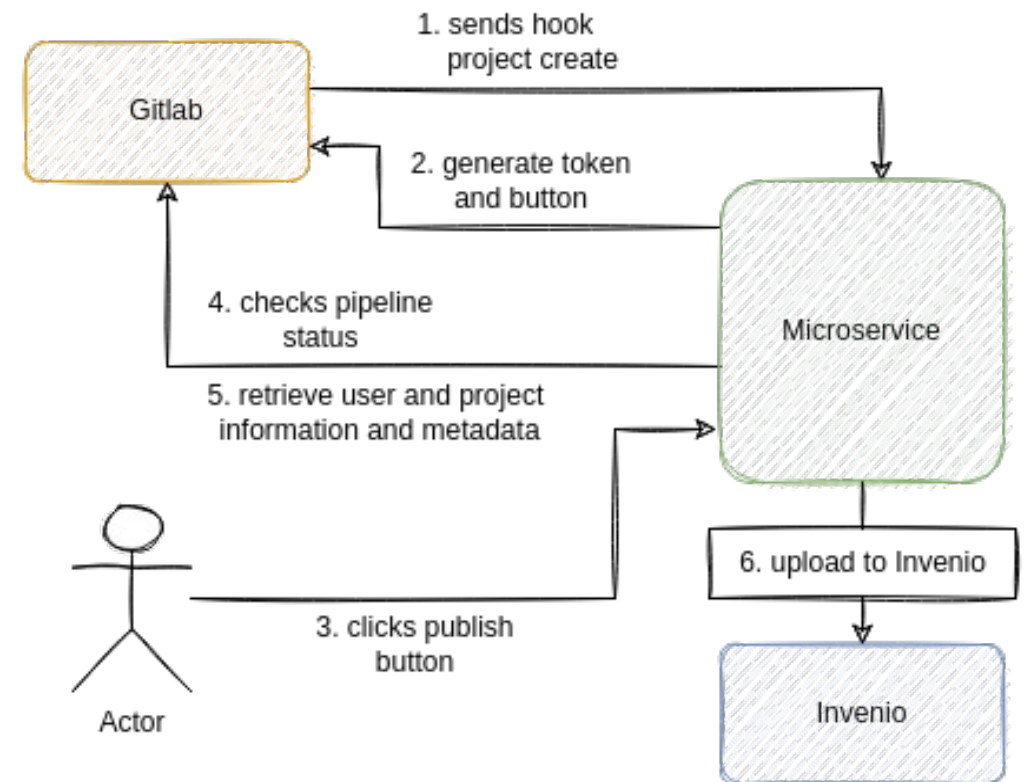
- Metadaten-Annotation-Tool
 - Annotation von Metadaten über Webservice
 - Übergabe an InvenioRDM via REST-API
- BinAC HPC-Cluster
 - Hook im Job-Scheduler
 - Automatische Drafts nach Abschluß des Rechenjobs

- Community-Freigabe-Workflows



Workflows mit InvenioRDM

- DataHUB in DataPLANT
 - Annotated Research Context (ARC) wird über GitLab-CI automatisch auf Vollständigkeit geprüft
 - Veröffentlichungsprozess wird per Knopfdruck gestartet
 - Microservice überprüft Status des CI-Jobs & erstellt einen “Submission Request” innerhalb der DataPLANT-Community
 - DataPLANT Kurator validiert
 - Publikation wird veröffentlicht und DOI wird vergeben



Ausblick

- Verpflichtung zu einer langfristigen Verfügbarkeit von Forschungsdaten:
 - Technischer Aufbau der Repositorien
 - Benötigt zukunftssicheres Betriebskonzept
- Entwicklung im Sinne von Open Data und Open Science
- Befördert Kulturwandel in der Wissenschaft, d.h. Daten (und Code etc.) werden zitierbar und Teil des Scholarly Records
- Aufbau einer InvenioRDM User Group via AK FDM o.ä.

Ausblick

- Entwicklung und Erprobung einer effizienten, föderierten Backend-Infrastruktur
 - Nutzung von bwSFSv2 für effizientes Erasure Coding in einem über vier Standorte verteilten Object Storage
 - Gemeinsames Repository-System für die vier Universitätsstandorte Tübingen, Freiburg, Stuttgart und Hohenheim
 - Erweiterte Integration der bwIDM(2) Infrastruktur für Rechte/Rollen

Thank you for your attention!

Data))((PLANT



UNIVERSITÄT HOHENHEIM

Universität Konstanz



EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



dkfz. DEUTSCHES
KREBSFORSCHUNGSZENTRUM
IN DER HELMHOLTZ-GEMEINSCHAFT



plant-biotech.net



MAX-PLANCK-INSTITUT
FÜR VERHALTENSBIOLOGIE

EMBL



European Molecular
Biology Laboratory



zbsa
zentrum für
biosystemanalyse



UNIVERSITÄTS
KLINIKUM
ulm

BioDATEN is
sponsored by:



Baden-Württemberg



MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST