

# **bwFDM-Communities – Wissenschaftliches Datenmanagement an den Universitäten Baden-Württembergs**



Forschungsdatenmanagement an den Universitäten des Landes Baden-Württemberg – Bestandsaufnahme und Empfehlungen

*Präsentation des Projektes „bwFDM-Communities“*

Achim Streit – Karlsruher Institut für Technologie (KIT)

# Motivation

- Wachsende Zahl an Wissenschaftsdisziplinen kann durch effizienten Umgang mit Forschungsdaten einen schnelleren Erkenntnisgewinn erzielen
  - Digitale Revolution, Informationszeitalter, „in silico“, Big Data, Smart Data
  - Hohe fachliche Spezialisierung der Disziplinen
- Intelligentes Datenmanagement (sowie Analyse)
  - Mehr und hochwertigere Veröffentlichungen → Publikationsdruck
  - Sinkende „Halbwertszeiten“ aktueller Erkenntnisse und Ergebnisse
- Beachtung des gesamten Daten Lebenszyklus
  - Erstellung/Erzeugung, Datennahme, Metadaten-Anreicherung, Analyse, Veröffentlichung, Archivierung
- Zunehmende Forderungen: „Wissen“ wiederverwendbar zu erzeugen (z.B. Horizon2020, GWK-Gesamtkonzept)



# Forschungsdatenmanagement (FDM)



## ■ Forschungsdaten

- DFG: „... digitale und elektronisch speicherbare Daten, die im Zuge eines wissenschaftlichen Vorhabens, zum Beispiel durch Quellenforschung, Experimente, Messungen, Erhebungen und Befragungen, entstanden sind...“

[http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung\\_forschungsdaten\\_1001.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_forschungsdaten_1001.pdf)

- Digitale Laborbücher, ge- und verteilte Datenbestände, Bilder (Mikroskopie, Schriften), Messreihen, Observationsdaten, Experimentdaten, Simulationsdaten, ...

## ■ Forschungsdatenmanagement

- Datenorganisation, Datenzugriff, Open Data
- Kuration und Preservation, Repositorien, Archivierung, Bit-Erhaltung
- Annotation mit administrativen und deskriptiven Metadaten
- Datensicherheit, Verschlüsselung, Datenschutz
- Migration, Transfer, Replikation, Föderierung
- Datenanalyse und visuelle Datenverarbeitung

# Projektidee



- Hilfestellung zur Beantwortung der Frage:
  - Welche Infrastrukturen, Dienste und Technologien im Umgang mit Forschungsdaten sind nötig (oder sind auszubauen), um Baden-Württemberg als Wissenschaftsstandort langfristig eine Spitzenposition zu sichern?
- Vorgehensweise:
  - Kontaktaufnahme zu möglichst allen wissenschaftlichen Forschungsgruppen („Long Tail of Science“)
    - Konkrete Bedarfserhebung in persönlichen Gesprächen
    - Erfassung bereits vorhandener Lösungen und Erzielung von „Quick Wins“
    - Identifikation von technischen und organisatorischen Lücken im FDM
  - Vorschläge für Handlungsempfehlungen, abgeleitet aus den Interviewinhalten
- **NICHT:**
  - Software-Entwicklung, Betrieb oder Beschaffung von Hardware/Software

**„Quick Wins“:** bwSync&Share, Chemotion, Digitalisierungsdienst U-HD, elektronisches Laborbuch, dedizierte Lösungen

# Projektteam

- Heterogene Ansiedlung in RZs und Bibliotheken unter Einbeziehung weiterer Personen (insgesamt 26 Personen)
- Zusammenstellung existierender Lösungen von RZs, BIBs und anderen Wissenschaftseinrichtungen (z.B. SFBs, GESIS)
- Persönliches Kontaktieren möglichst aller wissenschaftlichen Forschungsgruppen
- Einfach realisierbare Bedarfe werden an entsprechende Dienstbetreiber (RZ, BIB, etc.) weitergegeben und realisiert
- Weitere Bedarfe werden aggregiert erfasst und dann gemeinsam analysiert und dokumentiert



# Datenerhebung



- Zeitraum: Januar 2014 – November 2014
  - Kontaktaufnahme und Identifikation von Ansprechpartnern
  - Testphase zur Optimierung des Fragebogens
  - Hauptbefragungsphase mit 571 Interviews (713 befragten Personen)
    - Ausschöpfungsquote von ca. 40%
- Standardisierter Fragebogen/Interviewleitfaden
  - 27 Hauptfragen mit 13 Unterfragen  
[http://bwfdm.scc.kit.edu/downloads/bwFDM-Communities\\_Interviewleitfaden.pdf](http://bwfdm.scc.kit.edu/downloads/bwFDM-Communities_Interviewleitfaden.pdf)
  - Gesprächsdauer: 80 Minuten im Durchschnitt (min. 15 Min., max. 260 Min.)
- Auszug aus dem Interviewleitfaden
  - Forschungsgebiet, Größe der Forschungsgruppe, Gegenstand der Forschung
  - Umfeld der Daten: Quellen, Datentypen, Volumen, Personenbezug, Industriepartner, Formate
  - Prozess der Datenverarbeitung, Lagerung, Publikation und Data Sharing, Allgemeines

# Methodik der Datenaufarbeitung



- Kernaussagen der Interviews wurden in User Stories überführt → 2554 User Stories aus 571 Interviews
  - **Als** Politikwissenschaftler wünsche ich mir die Einrichtung eines geschlossenen Datenarchivs durch das Land Baden-Württemberg, um Kontrolle zu haben, wer Zugang zu meinen Daten bekommen kann.
  - **Als** Ingenieur in der Produktionstechnik wünsche ich mir ein einheitliches, universelles Tool, um analoge Mess- und Simulationsdaten mit Metadaten zu versehen und zu archivieren.
  - **Als** Zoologe wünsche ich mir einen institutseigenen Server, um die am Institut erzeugten Daten zentral speichern und archivieren zu können.
- Zuordnung zu Themen und Fachgebieten, gegenseitiger Review der User Stories mit Doubletten-Entfernung
- Abstraktion in Meta User Stories als Zusammenfassung
  - **Als** Forschende wünschen wir uns ein Speicherarchiv, welches einfach zu bedienen und ausfallsicher ist, sowie eine von mir bestimmbare Zugriffsbeschränkung bietet, um unsere Forschungsdaten zur Nachnutzung sichern zu können



# Tools



Home » formular tabelle karchivierung Engelgot als Achim Stred (Gruppe Tristram)

## Tabelle kArchivierung

VIEW

Datensatz 1 von 110

[Sichtbar im Formular](#) [Sichtbar in Tabelle](#)

[Daten exportieren](#)

genannterFachbereich	DFG_Fachbereich	DFG_Wissenschaftsbereich	Hintergrund	Wunsch	Zweck
4 7 8 2 F Linguistik	Sprachwissenschaften	Geistes und Sozialwissenschaftler	A Gruppenleiter	w ein geeignetes Repository für die Langzeitarchivierung meiner Forschungsdaten	die Forschungsdaten langfristig archiviert, zitierfähig und ohne Restriktionen
4 8 0 2 K Physikalische Chemie	Physikalische und Theor	Naturwissenschaften	a Wissenschaftler, der große Datenm	w zentralen Speicherplatz mit automatischer Datenabholung von unserem Server und c	meine Daten langfristig (> 10 Jahre) unterbringen zu können.
5 8 6 2 H Tierernährung und Tierernähr	Agrar , Forstwissensch	Lebenswissenschaften	A Wissenschaftler, der langfristige Archiv	w eine einfache Möglichkeit	unsere AG Daten (bis zu 100GB pro Jahr/Mitarbeiter) langfristig archi
5 8 1 2 K Bauingenieur	Bauwesen und Architekt	Ingenieurwissenschaften	A Wissenschaftler, welcher damit anfäng	w eine Lösung zur langfristigen (nach DFG ca. 10 Jahre) Speicherung meiner Videodat	auch später Zugriff auf diese Daten haben zu können.
5 9 1 2 H Tierernährung und Tierernähr	Agrar , Forstwissensch	Lebenswissenschaften	A Wissenschaftler, der langfristige Sicher	w eine einfache Möglichkeit, meine Daten (ca. 200 GB pro Jahr pro Mitarbeiter) langfris	später diese Daten durch neue leistungsfähigere Methoden erneut an
5 9 1 2 S Strömungsmechanik	Wärmeenergietechnik T	Ingenieurwissenschaften	A Wissenschaftler	w bei zentraler Datenhaltung eine klar nachvollziehbare Festlegung der Zugriffsrechte n	diese Datenhaltung akzeptieren zu können.
5 1 1 2 S Raum und Umweltforschung	Bauwesen und Architekt	Ingenieurwissenschaften	A Wissenschaftler mit personenbeziehb	w einen zugriffskontrollierten Archivspeicher am RZ mit einer Kapazität von ca. 2 TB (B	die langfristige Sicherung dort vollständig hin zu verlagern.
5 1 1 2 S Biomechanik	Medizin	Lebenswissenschaften	A Wissenschaftler	w Finanzierung von wesentlich umfangreicher Kapazitäten (50TB) zur langfristigen (z.B	die Nutzung solcher Daten für langfristige medizinische Fragestellung
5 1 1 2 F Geisteswissenschaft/Linguistik	Sprachwissenschaften	Geistes und Sozialwissenschaftler	A Wissenschaftler	w eine verlässliche und einfach handhabbare / anwendbare Möglichkeit zur Langzeitarc	meine Daten speichern zu können.
5 1 1 2 T Geschichtswissenschaft	Geistes und Sozialwissenschaft	Geistes und Sozialwissenschaftler	A Wissenschaftler eines kleinen Fachge	w nach bspw. Personen und Orten digital durchsuchbares Archivgut	dieses nach bestimmten Fragestellungen auswerten zu können.
5 1 1 2 T Empirische Kulturwissenschaft	Außereuropäische Spra	Geistes und Sozialwissenschaftler	A Wissenschaftler	w ein fachspezifisches Datenarchiv	Forschungsdaten abzulegen, anzusehen und zitieren zu können.
5 1 1 2 T Sportwissenschaft	Medizin	Lebenswissenschaften	A Wissenschaftler	w ein fachspezifisches Datenarchiv	Forschungsdaten abzulegen, anzusehen und zitieren zu können.
6 1 1 2 I Bioinformatik	Grundlagen der Biologie	Lebenswissenschaften	A Wissenschaftler mit Simulationsdaten	w eine Storage Lösung, die mitverfolgt, welche Daten gebraucht werden, und nicht gene	den vorhandenen Speicher effizient zu nutzen und den Schritt der Arc
6 1 1 2 H Geologie	Geologie und Paläontol	Naturwissenschaften	A Gruppenleiter mit begrenztem Budget	w eine kostenfreie Lösung für Datenpublikation und archivierung	die langfristige Verfügbarkeit der Daten nicht an eine dauerhafte Baza
6 1 1 2 H Astronomie	Astrophysik und Astronc	Naturwissenschaften	A Projektleiter, der für die Datenhaltung	w eine nachhaltige Finanzierungslösung für Langzeitarchivierung	nicht an teure Dienste für die LZA gebunden zu sein, die aktuell teilw
6 1 1 2 T Indologie	Sprachwissenschaften	Geistes und Sozialwissenschaftler	A Wissenschaftler	w ein offenes webbasiertes Datenarchiv mit Permalinks, in das laufend Daten überführt	Forschungsdaten ablegen, ansehen und zitieren zu können.
6 1 1 2 T Physikalischer Chemie	Physikalische und Theor	Naturwissenschaften	A Wissenschaftler	w eine Finanzierung der Langzeitarchivierung durch die Forschungsinstitution (Uni)	den Erhalt der archivierten Daten auch nach Ausscheiden des Einzelf
6 1 1 2 F Neurologie	Neurowissenschaft	Lebenswissenschaften	A Wissenschaftler im Bereich der Neurok	w ein unkompliziertes Werkzeug	nach Projektende die Daten bequem und einfach archivieren zu könn
6 1 1 2 F Neurologie	Neurowissenschaft	Lebenswissenschaften	A Wissenschaftler im Bereich der Neurok	w ein sicheres Archiv / Repository, welches einen granular abgestuften geschützten Zu	Daten zur Verfügung stellen zu können und gleichzeitig zu wissen we
6 1 1 2 F Neurologie	Neurowissenschaft	Lebenswissenschaften	A Wissenschaftler im Bereich der Neurok	w eine Plattform	meine Simulationen als komplettes (lauffähiges) System zur Verfüg
6 1 1 2 F Neurologie	Neurowissenschaft	Lebenswissenschaften	A Wissenschaftler im Bereich der Neurok	w ein Archiv für Primärdaten / Experimentdaten, bei dem die meisten Daten auch auf B	die kostspielig gewonnen Daten brauchbar zu sichern.
6 1 1 2 H Kunstgeschichte	Kunst , Musik , Theater	Geistes und Sozialwissenschaftler	A Doktorand mit einer php/sql Datenban	w einen zentralen Dienst für die funktionale Langzeitarchivierung und open access des	die Inhalte und die Funktionalität dauerhaft verfügbar und leicht zugr
7 1 1 2 U Medieninformatik	Informatik	Ingenieurwissenschaften	A Wissenschaftler	w eine Möglichkeit, meine Daten (personenbezogene Daten), über die von der DFG gef	diese bei Bedarf nachnutzen zu können.

1: 89 Personal;;IT-Personal;;Politikwissenschaft;Konstanz;•Als Politikwissenschaftler wünsche ich mir einen Experten, der technische Unterstützung bei der statistischen Datenbereinigung und der Dokumentation der Daten

2: 191 Softwarewünsche;Automatisierung;;Psychologie;Konstanz;•Als Wissenschaftler wünsche ich mir automatisches Preprocessing, dass aus MRT-Aufnahmen die Neuronen extrahiert , weil dies das Preprocessing der Da

3: 275 Softwarewünsche;Softwarewünsche;;Biologie;Karlsruhe;•Als Wissenschaftler, welcher sehr häufig mit Imagej arbeitet und damit Bilder analysiert wünsche ich mir eine Automatisierung der manuellen Schritte, welc

4: 371 Repository;Datenbanken;;Anorganische Chemie;Heidelberg;•Als Wissenschaftler wünsche ich mir eine skalierbare open access Moleküldatenbank, die effizient durchsucht werden kann, , um die für den jeweiligen

5: 490 schnellere Datenleitungen;;Informatik;Karlsruhe;•Als bildlicher Wissenschaftler, welcher regelmäßig Backups seiner Forschungsdaten anlegen möchte, wünsche ich mir eine stabile und schnelle Verbindung (Ein

6: 493 HPC;mehr Rechenleistung;;Elektrotechnik;Karlsruhe;•Als Wissenschaftler, welcher sehr viele Simulationen durchführt, wünsche ich mir Zugriff auf Computer mit höheren Rechenleistungen als die aktuell verfügbar

7: 853 HPC;instituteigenen Rechencluster/lokalen Rechencluster;;Physik;Ulm;•Als Wissenschaftler der Rechencluster nutzt, wünsche ich mir - als Ergänzung zu den zentralen Rechenclustern (wie dem bwUniCluster) - ein

8: 874 Clouddienste;Clouddienste;;Wirtschaftsinformatik;Hohenheim;•Als Wissenschaftler, der regelmäßig Simulationen durchführt, wünsche ich mir ein benutzerfreundliches, flexibles Cloudsystem von virtuellen Servern,

9: 999 Softwarewünsche;Softwarewünsche;;Bauingenieur;Karlsruhe;•Als Wissenschaftler, welcher sehr viele Präsentationen (auch für Arbeitskollegen) vorbereiten muss und in der Lehre tätig ist wünsche ich mir eine visu

Anzahl der Datensätze: 46; keyWord: schnell

0: 33 HPC;lokalen Rechencluster;Processing/Erhebung;Physik (2);Konstanz;  
Als Physiker wünsche ich mir lokale Rechencluster am Speicherort meiner Daten um Simulationsdaten schnell durchrechnen und erhalten zu können.;1;1

1: 89 Personal;;IT-Personal;;Politikwissenschaft;Konstanz;  
Als Politikwissenschaftler wünsche ich mir einen Experten, der technische Unterstützung bei der statistischen Datenbereinigung und der Dokumentation der Datenanalyse leistet , um unsere Datensätze schneller und effektiver veröffentlichen zu können.;1;1

2: 191 Softwarewünsche;Automatisierung;;Psychologie;Konstanz;  
Als Wissenschaftler wünsche ich mir automatisches Preprocessing, dass aus MRT-Aufnahmen die Neuronen extrahiert , weil dies das Preprocessing der Daten erheblich verschleunern würd





# Außendarstellung

- Erfahrungsaustausch mit anderen Initiativen in D und EU
  - Humboldt-Universität (FDM policy und Umfrage)
  - DINI/nestor AG Forschungsdaten
  - Leibniz-Institut für Informationsinfrastruktur (FIZ)
  - Auf zahlreichen Workshops und Konferenzen mit Verantwortlichen bei DFG und Helmholtz, mit dem DCC (UK) und verwandten Projekten aus Österreich und der Schweiz
- Außendarstellung über
  - Newsletter, Webauftritte an Unis, Bekanntmachung in diversen lokalen Periodika der Unis, Twitter (2000 Follower), LSDMA, SDIL, re3data.org, BW-Wassernetzwerk, chemotion, Bio-Imaging Netzwerk, ...
- Mit Projektende online verfügbar
  - Abschlussbericht
  - User Stories und ausgewertete Interviewinhalte (anonymisiert)



# Newsletter

**bwFDM-Communities Newsletter**  
Veröffentlichung von Ergebnissen  
Juni 2015

**Diskutieren Sie mit!**

Wir möchten allen beteiligten Akteuren noch die Möglichkeit zur direkten Diskussteilnahme geben. Dazu haben wir auf einem Diskussionsportal [1] eine kurze Auswahl unserer Beobachtungen zusammen mit dem Idealzustand des Forschungsdatenmanagements veröffentlicht. Jeder Interessierte kann dort mit uns diskutieren und Ideen für konkrete Wege formulieren, wie man vom Ist-Zustand zu diesen Zielen kommt. In den nächsten zwei Wochen können Sie so eigene Ideen einbringen, die dann auch noch Eingang in unseren Abschlussbericht finden können.

**Erstes Projektmaterial online**

Wir möchten im Sinne der Nachvollziehbarkeit unser Datenmaterial online stellen. Den Anfang macht unser viel nachgefragter Interviewleitfaden, mit dem wir in die Gespräche mit den Forschenden gegangen sind und eine Aufstellung der Disziplinen mit denen wir geredet haben. Beides ist nun auf unserer Webseite online einsehbar. Unser „Datenschatz“ aus User-Stories folgt voraussichtlich bis zum 17. Juli.

**InfoBox**  
Zeitleiste für die nächsten Monate

- 15. Juni – Ende der aktuellen E-Science Ausschreibung [2] zum Forschungsdatenmanagement
- 30. Juni – Unser Ergebnisbericht wird an das MWK übergeben.
- 17. Juli – Abschlussveranstaltung für die Leitungsebenen der Universitäten in Baden-Württemberg, deren zentralen Einrichtungen sowie der externen Gutachter der E-Science Ausschreibung [2] in Stuttgart
- 30. September – Öffentliche Vorstellung der bwFDM-Communities Ergebnisse und der geförderten E-Science Projekte in Karlsruhe
- 30. September – voraussichtliches Ende der finalen Datenaufbereitungsphase und Online-Bereitstellung aller aufbereiteten Projektdaten von bwFDM-Communities

Noch ohne uns bekannten Termin:  
E-Science Ausschreibung zum Thema virtuelle Forschungsumgebungen

Webseite: <http://bwfdm.scc.kit.edu/>  
Liste der Ansprechpartner: <http://bwfdm.scc.kit.edu/mitarbeiter>  
Redaktion: [Frank.Trostman@kit.edu](mailto:Frank.Trostman@kit.edu)

[1] <https://www.dscadb.koelr.com/consultation/10042>  
[2] [https://www.kit.edu/bwfdm/redaktion/mwktroster/bwfdm/aktuelle\\_Ausschreibung/E\\_Science/E\\_Science\\_Ausschreibung\\_2015-endg.pdf](https://www.kit.edu/bwfdm/redaktion/mwktroster/bwfdm/aktuelle_Ausschreibung/E_Science/E_Science_Ausschreibung_2015-endg.pdf)

**bwFDM-Communities Newsletter**  
Juni 2014

**bwFDM-Communities Newsletter**  
August 2014

**bwFDM-Communities Newsletter**  
Oktober 2014

**bwFDM-Communities Newsletter**  
Dezember 2014

**Communities Newsletter**  
März 2015

**InfoBox**  
Gibon, genannt?

Der beliebteste Dienst bwSync&Share hat seine Online-Speicherkapazität pro Universitätsmitglied auf 25GB erhöht. Die ehemalige US-Sicherheitsberaterin Condoleezza Rice sitzt im Verwaltungsrat von Dropbox.

Workshops für disziplinübergreifende Interessensgruppen

LSDMA Community Forum 2015  
Thema: Large Scale Data Management and Analysis - Community Berichte  
Zeit und Ort: 25. - 26. März, Berlin

RDA-@HNC-Workshop 2015  
Thema: Datenmodelle und Standards  
Zeit und Ort: 28. - 29. Mai, Karlsruhe

DATA 2015  
Thema: Datenmanagement Technologien und Anwendungen  
Zeit und Ort: 20. - 22. Juli, Colmar (Elsass)

<http://bwfdm.scc.kit.edu/> (Webseite)  
<http://bwfdm.scc.kit.edu/mitarbeiter> (Liste der Ansprechpartner)  
[Frank.Trostman@kit.edu](mailto:Frank.Trostman@kit.edu) (Redaktion)



~1000 Empfänger

# Und nun zu den Ergebnissen...

