

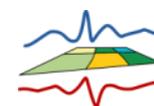
# Langzeitarchivierung aus der Perspektive eines interdisziplinären Großprojektes: Ein Erfahrungsbericht aus dem SFB/TRR 32

Einleitung – SFB/TRR 32 – Anforderungen – System  
Architektur – Herausforderungen  
& Probleme – Zusammenfassung

**Constanze Curdt**<sup>1,2</sup>, **Dirk Hoffmeister**<sup>1</sup>, **Georg Bareth**<sup>1</sup>, **Ulrich Lang**<sup>2</sup>

<sup>1</sup> Geographisches Institut, Universität zu Köln

<sup>2</sup> Regionales Rechenzentrum (RRZK), Universität zu Köln



# FDM in Großforschungsprojekten

## ■ Gründe für FDM

Reproduzierbarkeit der Forschungsergebnisse

Nachnutzung der Forschungsdaten

Austausch von Forschungsdaten zur Schaffung von Synergien

## ■ Forschungsdaten sollten...

gut aufbereitet, organisiert, dokumentiert, nachhaltig gespeichert und gesichert werden

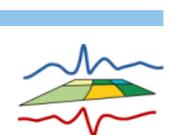
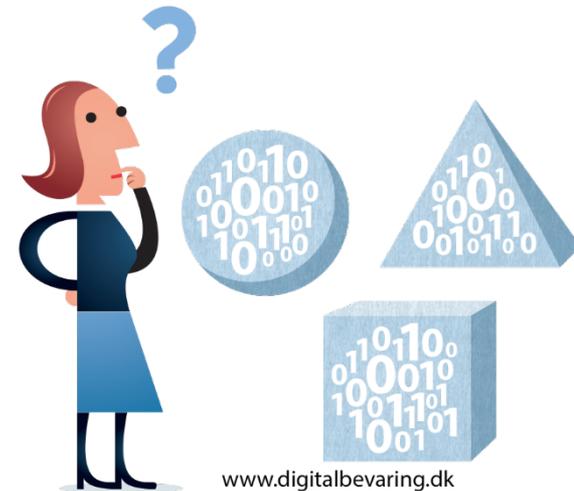
## ■ Merkmale von FDM

zeitlich begrenzte Projektlaufzeit und Finanzierung

Datenverfügbarkeit während und nach Projektlaufzeit

größtenteils projektspezifische FDM Lösungen

INF-Projekt in SFB oder SFB/TRR



# SFB/TRR 32

- „Muster und Strukturen in Boden-Pflanzen-Atmosphäre-Systemen: Erfassung, Modellierung und Datenassimilation“

finanziert durch DFG

Projektlaufzeit 12 Jahre (2007 – 2018)

- **Involvierte Standorte**

Universität Aachen, Bonn, Köln,  
Forschungszentrum Jülich

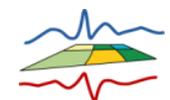
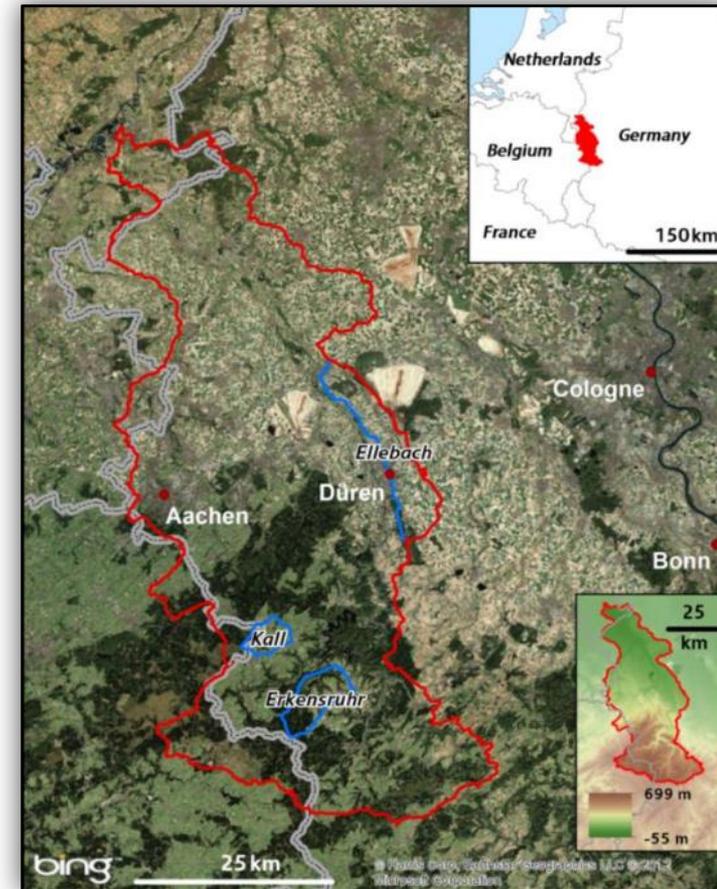
- **Verschiedene Forschungsdisziplinen:**

Geophysik, Bodenkunde, Pflanzenkunde,  
Hydrologe, Geoinformatik, Geographie,  
Fernerkundung, Meteorologie, Mathematik

- **Forschungsziel:**

Erforschung von Wechselwirkungen von  
Mustern und Strukturen im Boden-Pflanzen-  
Atmosphärensystem

- **Projektgebiet: Einzugsgebiet der Rur**



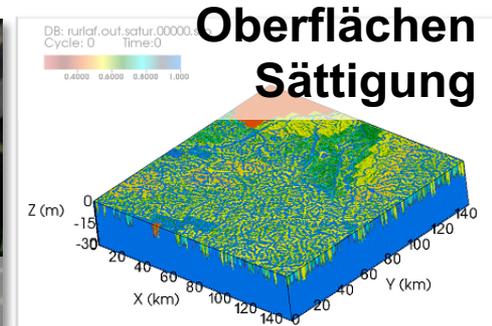
# SFB/TRR 32 – Projekt Daten



e.g. Temperatur, Druck, Wind



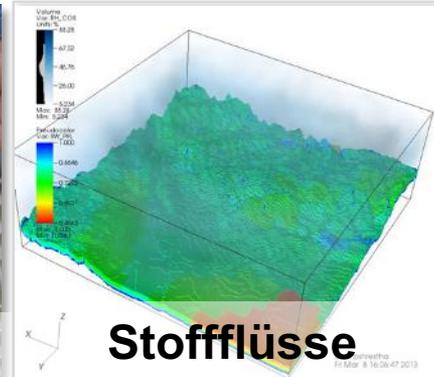
Blattmessungen



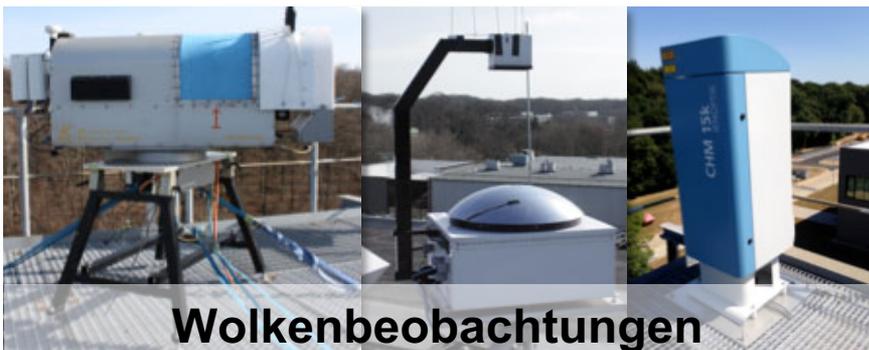
Hyperspectralmessungen



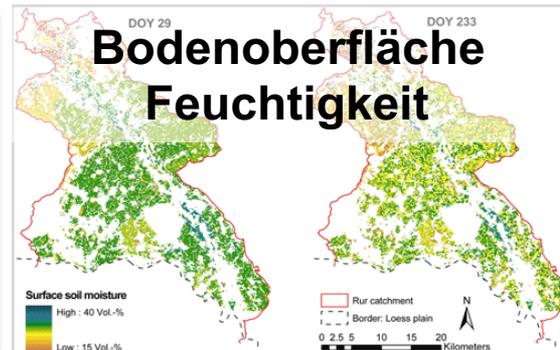
Pflanzenproben



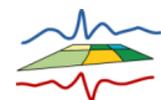
Eddy Kovarianz



Wolkenbeobachtungen

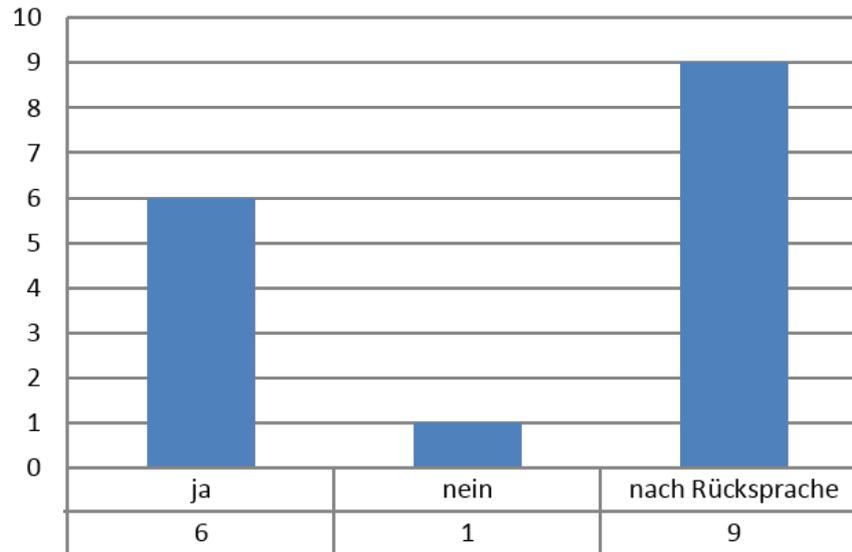


(TR32 2014)



# Anforderungen an das FDM im SFB/TRR 32

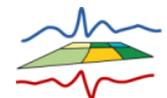
## 1.3 Wird Ihr TP Daten für die Projektdatenbank bereitstellen?



### ▪ Welche Art von Daten produzieren Sie und in welchem Format? Welche Größe haben Ihre produzierten Daten voraussichtlich?

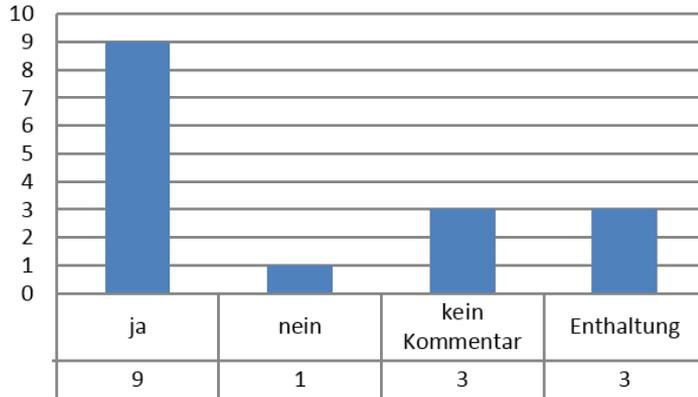
GRIB-Code(binär), NetCDF, binäres Datenformat (IEEE Standard), HDF5, ASCII, EXCEL, GIS und Fernerkungsformate, Text, ACCESS, ArcGIS Vektor und Raster Format, FAP Format, Binärformat

Datengröße pro Projektphase / Teilprojekt: <100MB bis 400GB

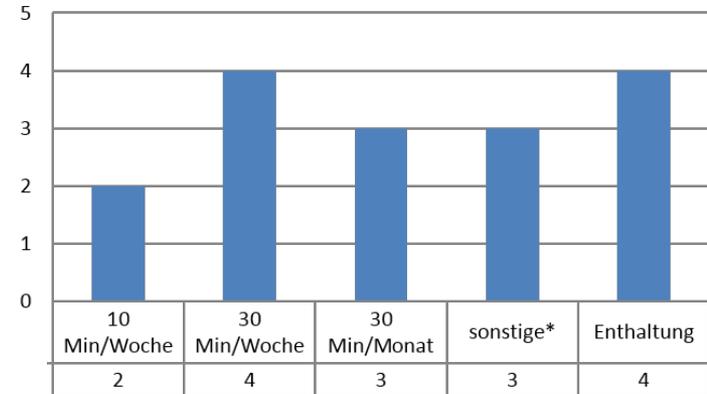


# Anforderungen an das FDM im SFB/TRR 32

## 1.6 oder ist die Einbindung von aggregierten und selektierten Daten sinnvoller?

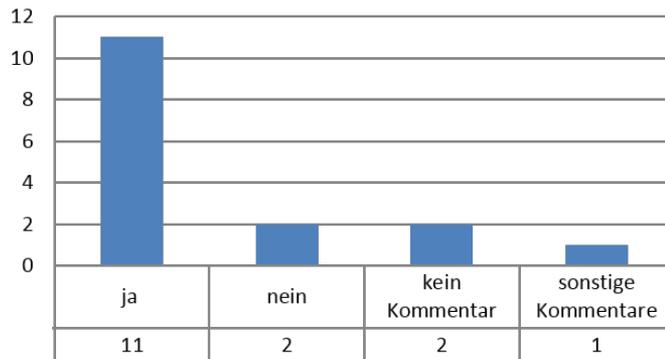


## 2.3 Wie viel Zeit können Sie für die Datenaufbereitung und -pflege aufwenden?

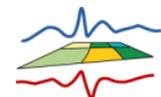
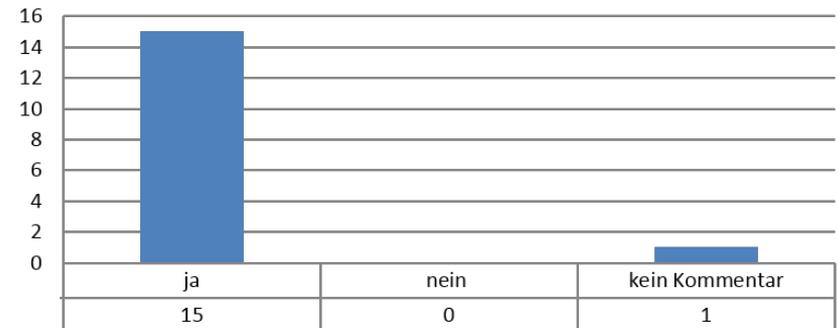


\* 180 Min/Woche, 240 Min/Woche, 60 Min/Monat

## 4.1 Unterstützen Sie eine Projektdatenbank, auf die über das Internet (nur für Benutzer mit Passwort) zugegriffen werden kann?



## 4.4 Finden Sie es wünschenswert, dass auf zusätzliche Informationen (z.B. Publikationen, Dissertationen, Berichte etc.) in der Projektdatenbank zugegriffen werden kann?



# Anforderungen an das FDM im SFB/TRR 32

## ▪ Anzahl an Daten

Projektlaufzeit 12 Jahre (2007-2018)

ca. 65 Teilprojekte → ~ 200 Wissenschaftler

## ▪ Datenvielfalt:

heterogene Daten in verschiedenen zeitlichen und räumlichen Skalen

Forschungsdaten, Publikation, Berichte, Karten, Bilder, ...

disziplinspezifische Dateiformate

unterschiedliche Dateigrößen (von kB bis GB)

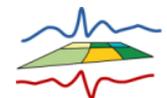
## ▪ Empfehlungen der DFG:

FDM in Kooperation mit Bibliothek oder Rechenzentrum

bestehende Infrastruktur und Techniken nachnutzen

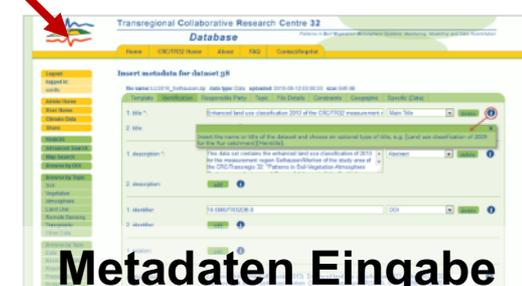
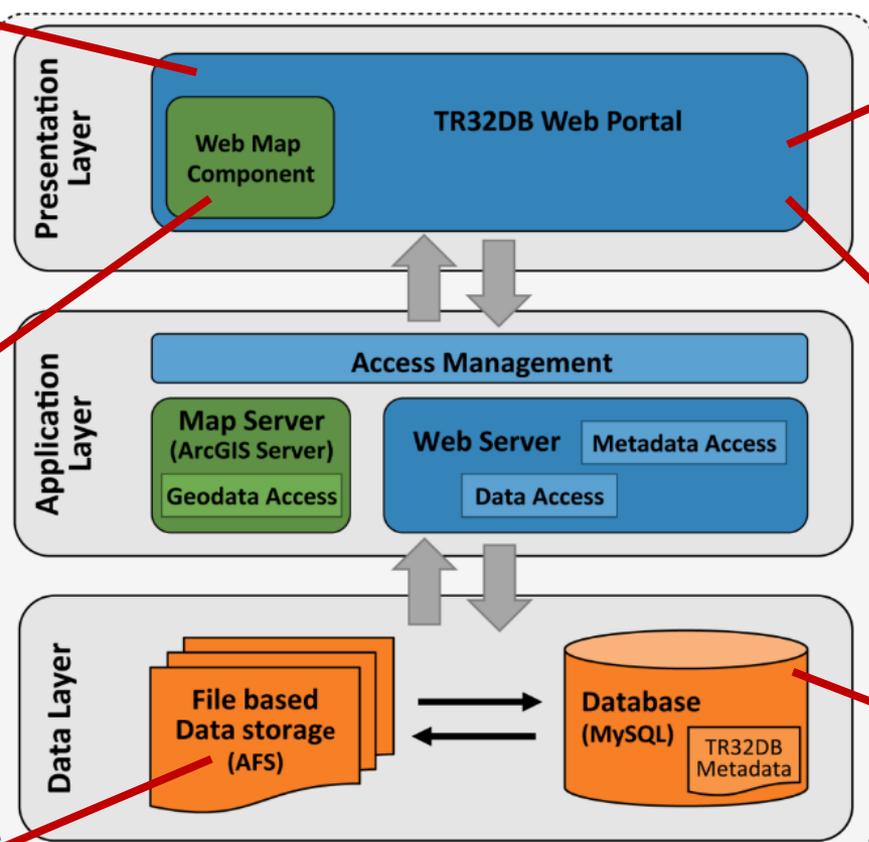
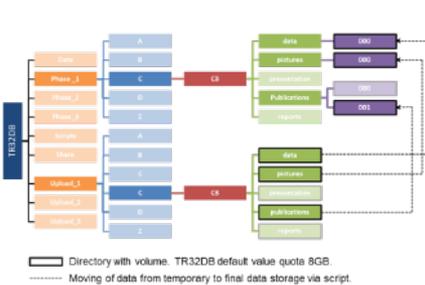
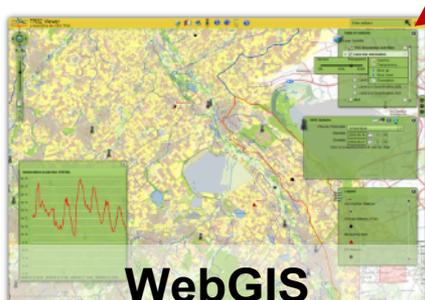
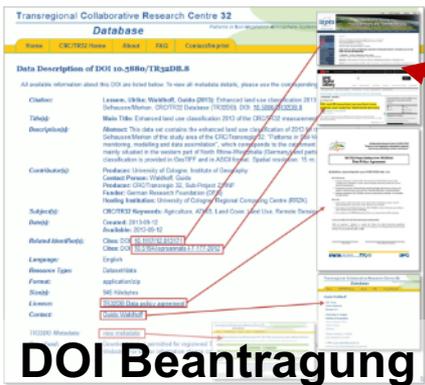
Datenverfügbarkeit über Projektende hinaus

Gute Wissenschaftliche Praxis



# SFB/TRR 32 Projekt Datenbank (www.TR32DB.de)

## Design und Implementierung einer 3-Tier-Architektur



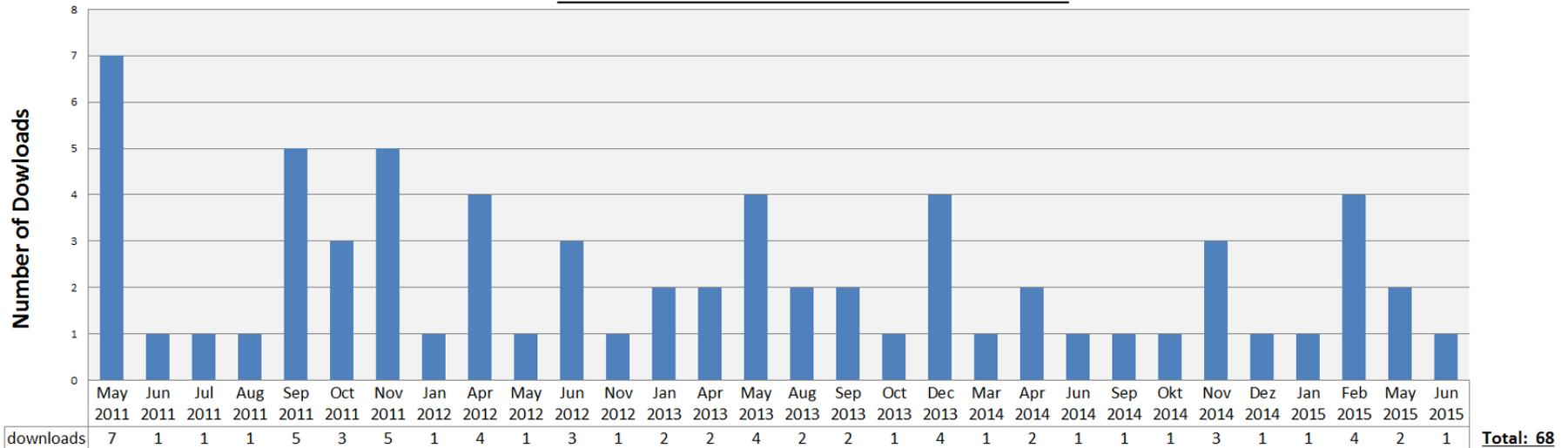
\* hosted at Regional Computing Centre Cologne (RRZK)



# Herausforderungen und Probleme

## Download of Land Use Classification Data via TR32DB

### Download of Dataset ID 211\* via TR32DB



\* SEIBERT, S., AUERSWALD, K., FIENER, P., DISSE, M., MARTIN, W., HAIDER, J., MICHAEL, A., GERLINGER, K., 2011. Surface runoff from arable land - a homogenized data base of 726 rainfall simulation experiments, CRC/TR32 Database (TR32DB). doi: 10.1594/GFZ.TR32.2.

(Period: 11 May 2011 - 17 June 2015)

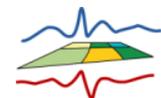
## Project Sections

(Period: 3 May 2011 - 23 February 2015)

## ■ TR32DB Daten (Stand 17. Juni 2015)

~370 GB Data in AFS Datenspeicher

~1100 Metadatenätze



# Zusammenfassung



## ■ Projektdatenbanken

limitierte Lebensdauer durch Finanzierung  
bevorzugt projektspezifische Lösungen

- Zusammenarbeit mit Bibliotheken, Rechenzentren, ...
- vorhandene Infrastruktur verwenden
- Systemaufbau in enger Zusammenarbeit mit Forschern

## ■ Forschungsdaten

heterogene, disziplin-spezifische Datenformate und -größen

- interoperable Metadaten, Datendokumentation
- Datenqualität ist wichtig

## ■ Umgang mit Forschern

- Zusammenarbeit, frühzeitige Einbindung
- Beachtung von Anforderungen
- viel Training und Support
- Geduld und Verständnis



www.digitalbevaring.dk

