

# Forschungsdatenmanagement am Helmholtz-Zentrum Dresden-Rossendorf und am Helmholtz-Zentrum Berlin (RDM@DB)

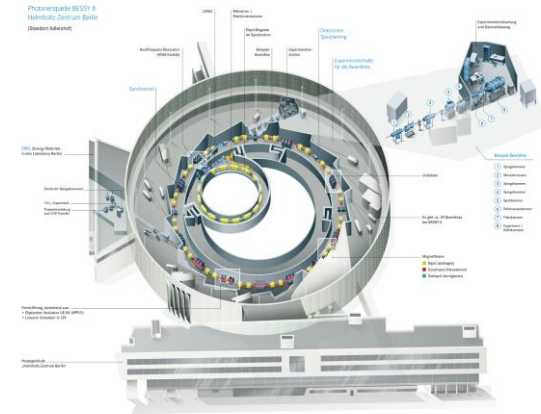
- Ausgangspunkt und Ziele des Projektes
- Arbeitstand
- Nachhaltigkeit

Dr. Uwe Konrad, Heike Görzig, Dr. Guido Juckeland  
Berlin 18.10.2018  
DOI: 10.14278/rodare.62



# Ausgangspunkt

- HZDR and HZB waren im Helmholtz Forschungsbereich „Materie“ mit neuen Generationen von Detektoren an den Beamlines konfrontiert
  - sehr hohe Datenraten ( $\sim 50$  GB/s)
  - dezentrales Datenmanagement nicht mehr mögl.
- Forschungsdaten-Richtlinie am HZB war vorhanden, am HZDR in Vorbereitung
- Gemeinsam entwickelte Software „GATE“ zur Begleitung der organisatorischen Abläufe
- Forschungsdaten-Management-Plattformen waren in Entwicklung



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

**HZB** Helmholtz  
Zentrum Berlin

**DRESDEN**  
concept

**HZDR**

# Ziele

Entwicklung von Forschungsdaten-Management-Lösungen (FDM) mit folgenden Eigenschaften:

- offen, generisch und mit Publikationsfunktion (DOI)
- Ablage auch sehr großer Datensätze
- Eignung für verschiedene Communities
- open-source, on-premise

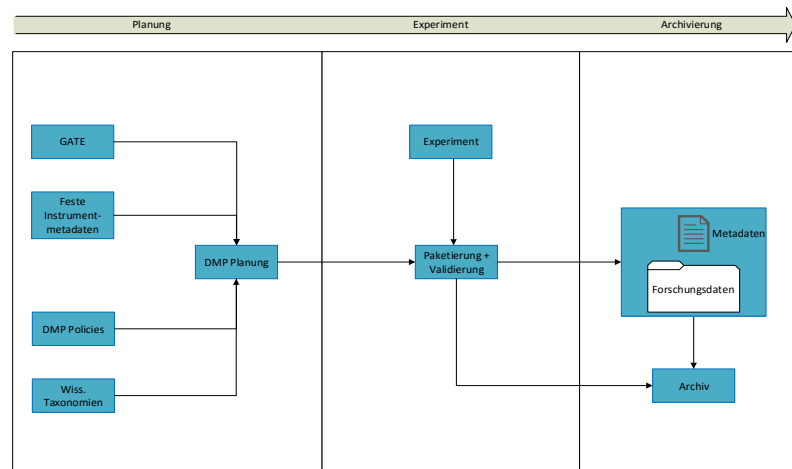
Die FDM-Lösungen sollen in den Nutzerbetrieb der Großgeräte integriert, aber auch von den Forschern der Institute verwendet werden.

- Verlinkung zu Publikationssystemen
- Verlinkung zu Software-Repositories
- Abgestimmte Data Policy

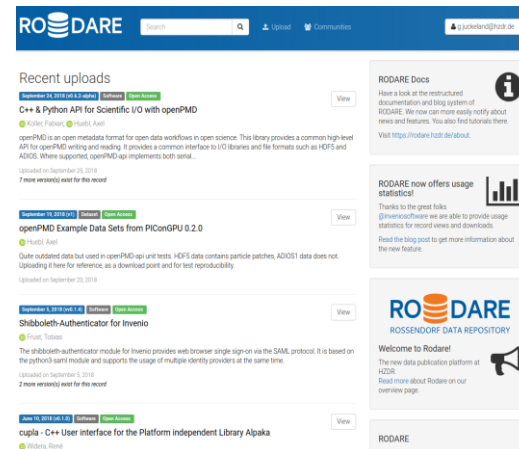
GEFÖRDERT VOM

# Arbeitsstand - Arbeitspakete

- AP1: Analyse der Datenflüsse + Workflows
- AP2: Entwicklung der Datenmodelle
- AP3: Test/Auswahl vorhandener Lösungen
- AP4: Pilotsystem
- AP5: Verknüpfung mit vorhandenen Systemen



GEFÖRDERT VOM



# 1. Analyse: Interviews

- Interviews mit den Forschergruppen von Experimenten unterschiedlicher Komplexitäten und Datenmengen aus verschiedenen Forschungsbereichen
- Interviews mit den Betreibern der begleitenden Infrastrukturen und Software-Werkzeuge
- Diskussion und Abstimmung in der Community der europäischen Lichtquellen(PANdata, Calypso+, LEAPS) und der Research Data Alliance

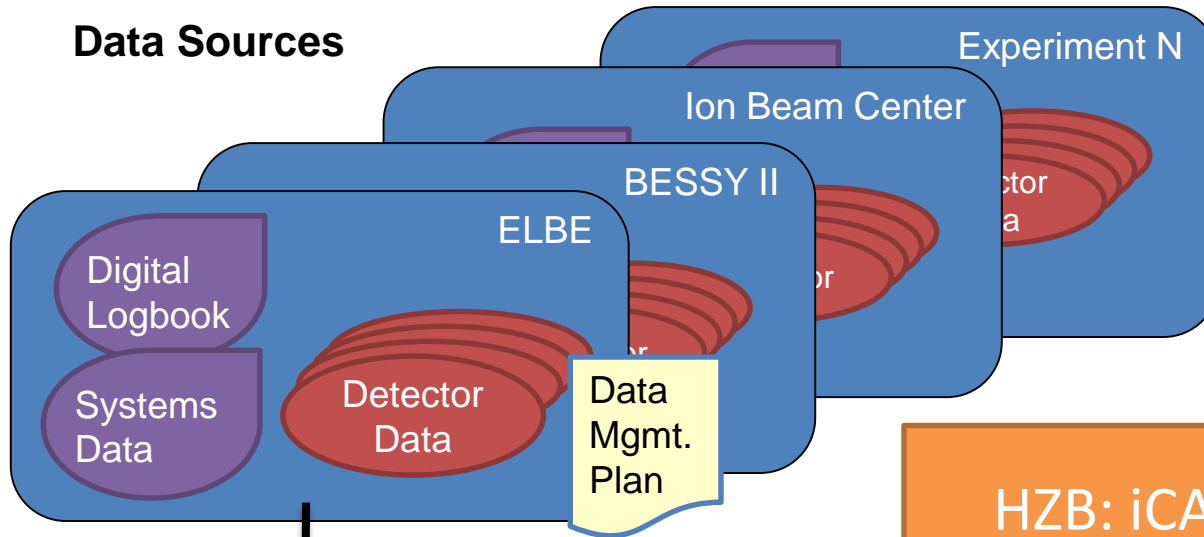
→ sehr uneinheitliche, nicht standardisierte Metadaten an den Instrumenten



GEFÖRDERT VOM

# Analyse: Workflows und Werkzeuge

## Data Sources



## EU / HGF Projects:

- PaNData,
- CALIPSO+
- LEAPS
- HGF Inkubator

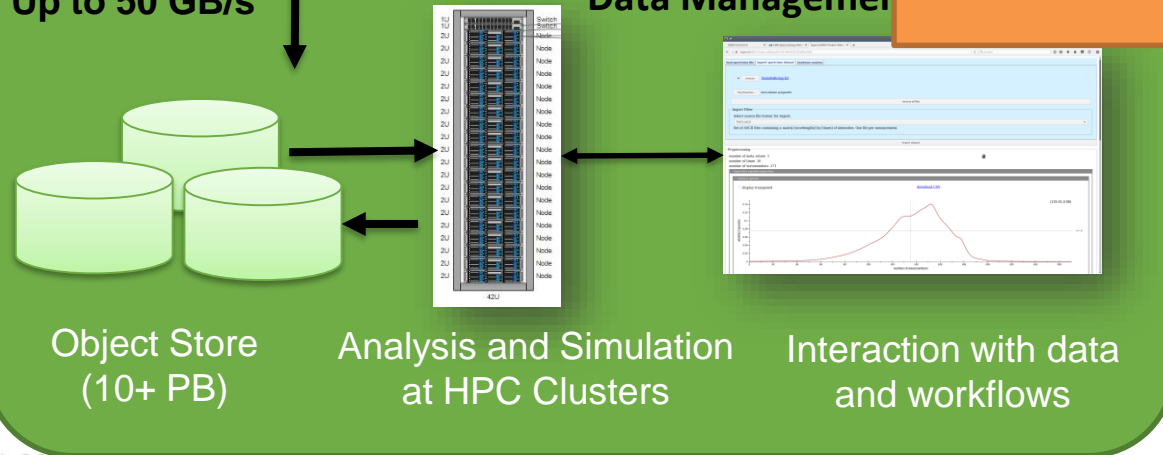
Data  
Policy

Public.  
Policy

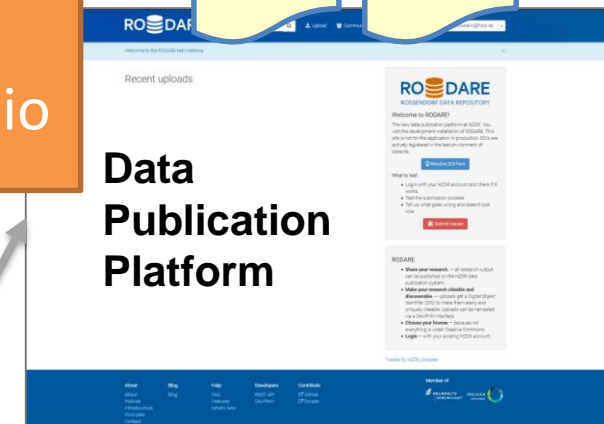
HZB: iCAT  
HZDR: Invenio

Up to 50 GB/s

Data Manager



Data  
Publication  
Platform



Long  
term  
Archive

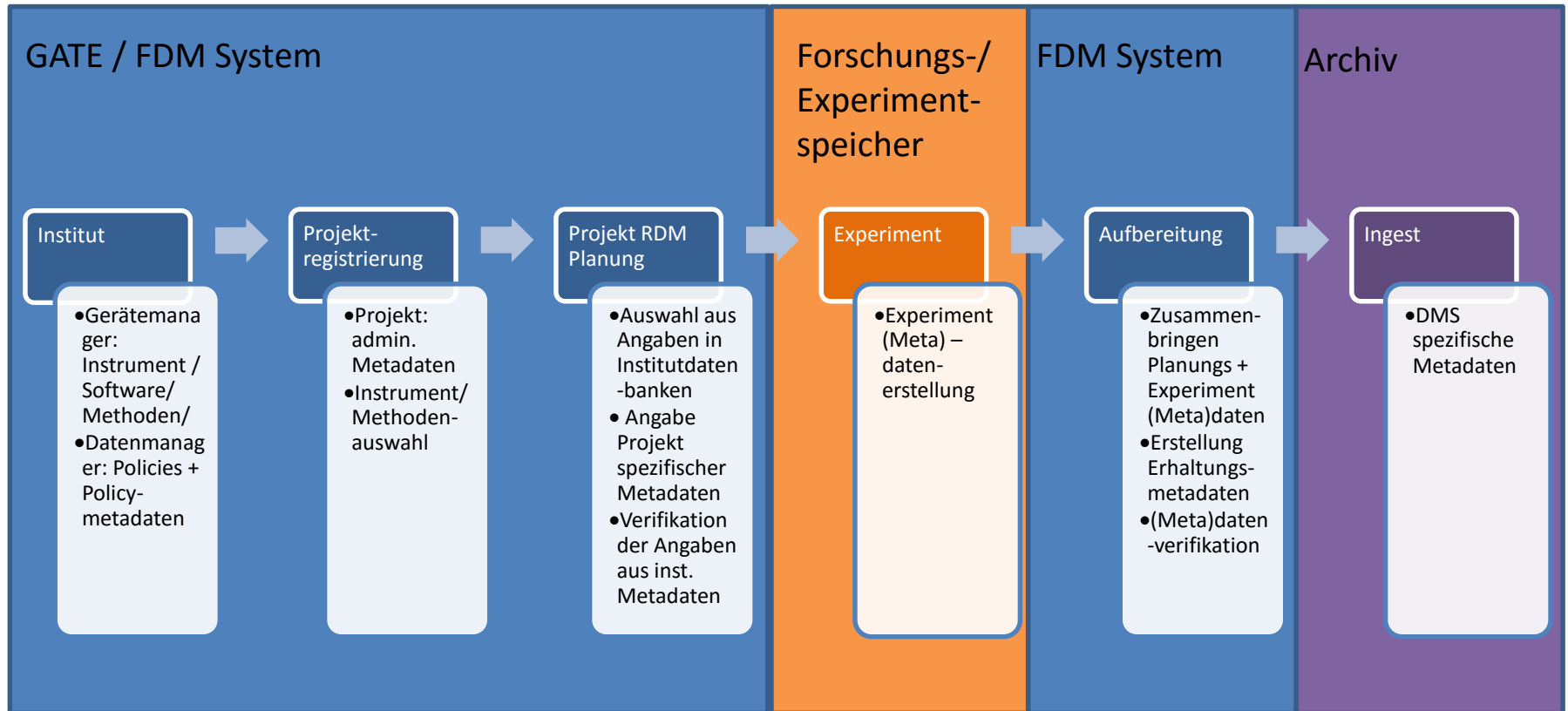
Partner  
WWW

HZB  
Helmholtz  
Zentrum Berlin

concept

Member of the Helmholtz Association

# FDM Datenfluss und Speicher



GEFÖRDERT VOM

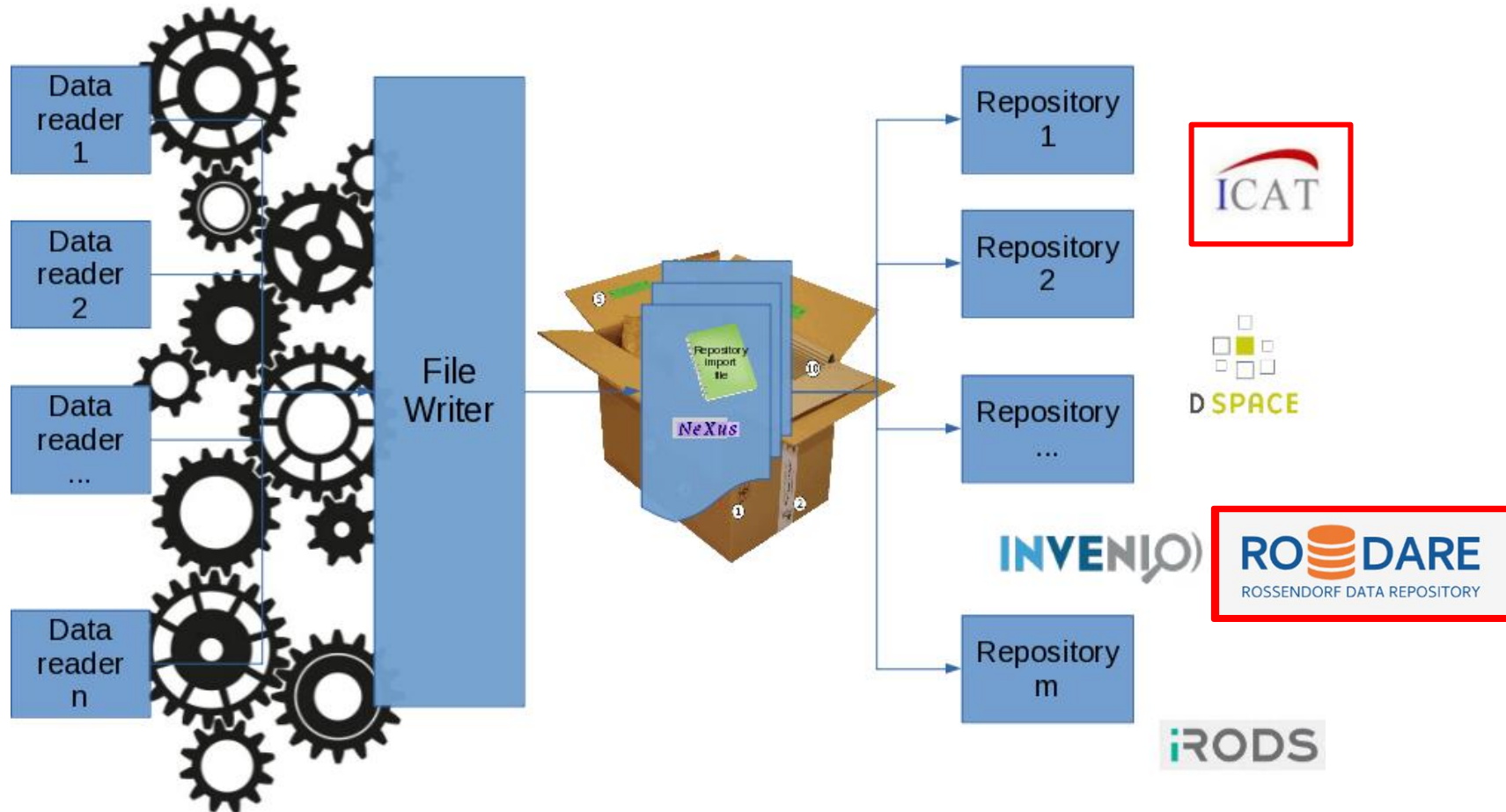
## 2. Modelle

- Metadatenmodell
  - Instrument/Methode haben einen Kern statischer Metadaten
  - Projektspezifische Metadaten
  - Experimentspezifische Metadaten
- Geschachteltes Datenmodell
  - DataCite Mindestkriterien
  - Community spezifische Meta-Daten (Anlagen-spezifische Daten, Laborbuch-Notizen)
  - Erhaltungsmetadaten
- Verarbeitungsmodell
  - Erhalt statischer und projektspezifischer Metadaten einheitlich (Nutzerportal GATE, Instrumentendatenbank, Publikationsworkflow)
  - Experiment ist anlagenspezifisch, benötigt ein flexibles Workflow-Management

GEFÖRDERT VOM



### 3. Lösungen: Community Formate für Repositorien



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

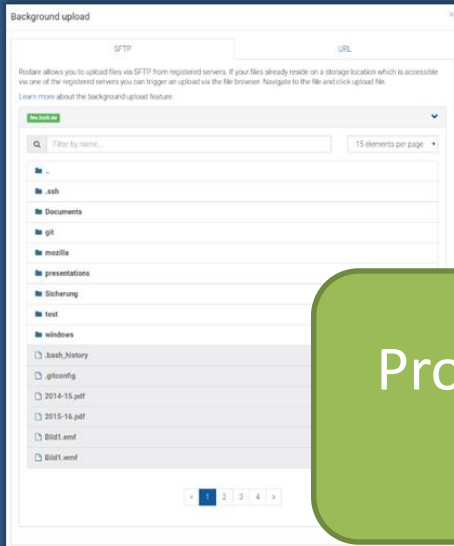
HZB  
Helmholtz  
Zentrum Berlin

DRESDEN  
concept

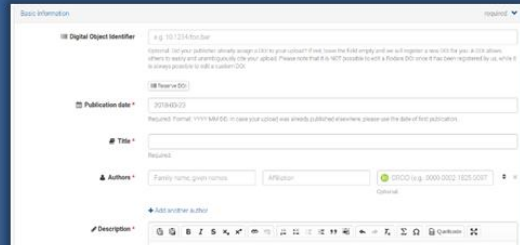
HZDR

# Nutzer Workflow

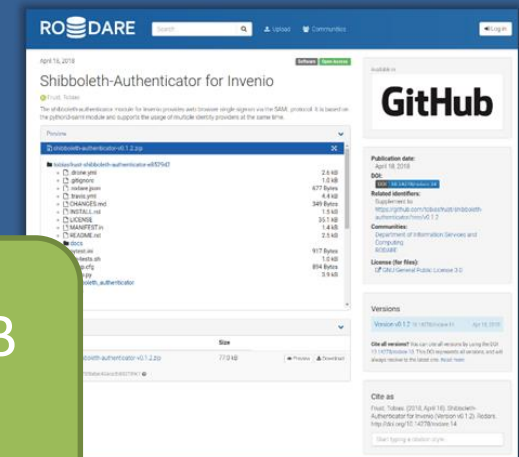
## Upload



## Describe



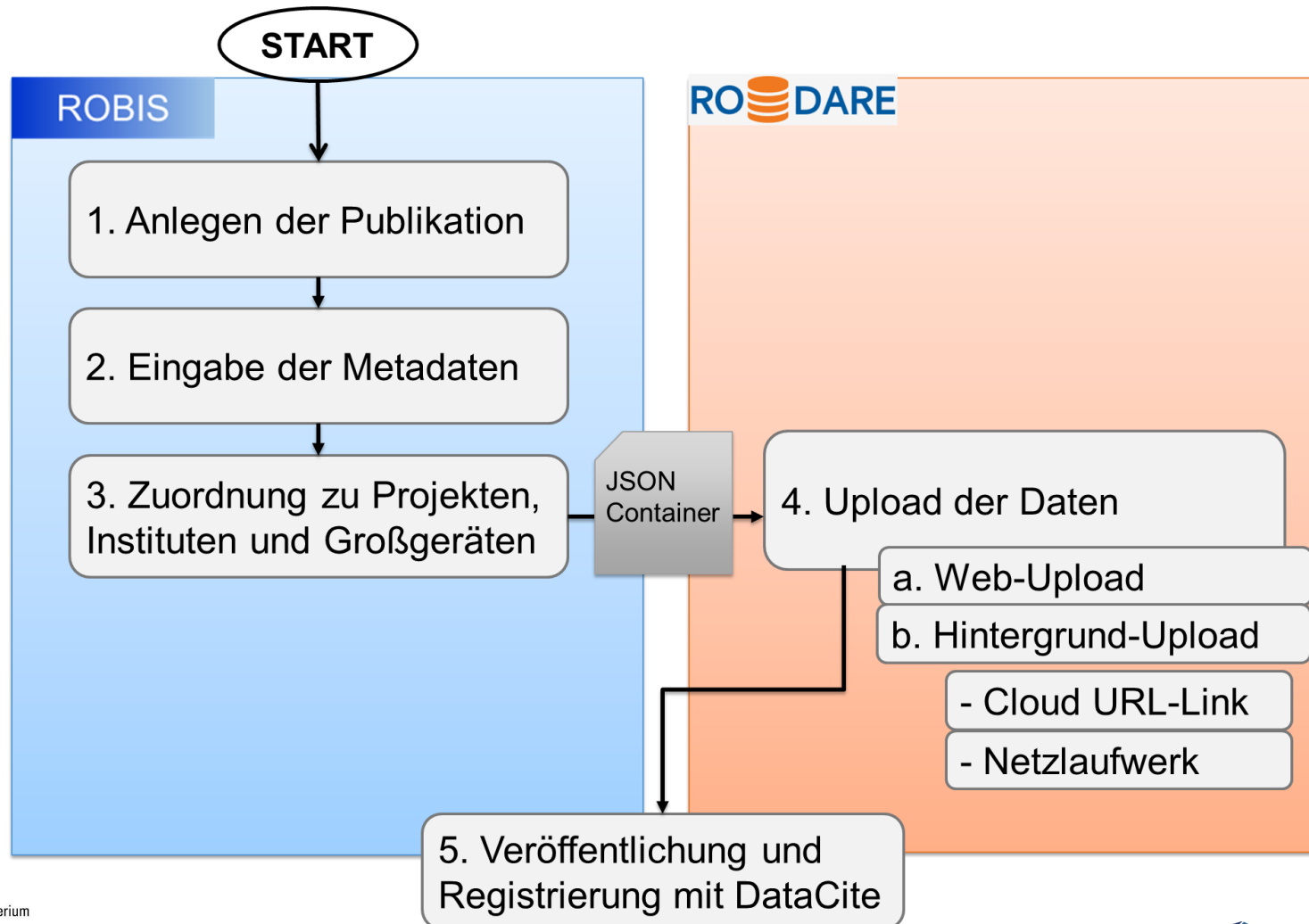
## Publish



## Prozesse im Rahmen von RDMatDB laufen z.T. auch automatisiert

GEFÖRDERT VOM

# Kopplung mit Publikationsdatenbank am HZDR



GEFÖRDERT VOM

# Nachhaltigkeit

## Wie erfolgt die langfristige und institutionenübergreifende Sicherstellung der Verfahren?

- Gemeinsam Entwicklung mehrerer Einrichtungen von Beginn an
- Intensive Abstimmung in internationales Gremien (PanDATA, LEAPS, RDA)
- Verwendung erfolgreicher open source Lösungen
- Betrieb der Infrastrukturen gemeinsam durch die Rechenzentren und Bibliotheken der Forschungseinrichtungen

## Wie können die erzielten Ergebnisse und Lösungen auf Dauer etabliert werden?

- Verankerung in bindenden Richtlinien (Data Policies) für Großgerätenutzer und Forscher
- Strikte Registrierung mit weltweit eindeutiger Identifikation (DOI)
- Kopplung von Ergebnis-, Daten- und Softwarepublikationen

GEFÖRDERT VOM

# Projekt RDM@DB

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

Vielen Dank für die Unterstützung  
und Ihre Aufmerksamkeit!

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

**HZB** Helmholtz  
Zentrum Berlin

DRESDEN  
concept

**HZDR**