

## RDM Workshop - Archiving and publishing your data at UFZ (2023)

Paul Remmler, Janine Felden (PANGAEA) – 07.07.2023



# Your RDM Team at UFZ

---

Our main strands:

- development of data infrastructures, services and tools
- scientific data processing / data pipelines
- consulting, training and networking



Intranet: <https://www.intranet.ufz.de/rdm>

Mattermost: <https://mm.ufz.de/ufz/channels/rdm--wkdv-news>

Mail: [rdm-contact@ufz.de](mailto:rdm-contact@ufz.de)



# Workshop agenda

---

1. **Welcoming / getting to know each other / orientation**
2. Data publication in general
3. Using UFZ- Data Management Portal (DMP) – Our Archive System
4. Using PANGAEA Data Publisher
5. Wrap-up and outlook

# Welcome to workshop Archiving and Publishing at UFZ

---

Using collaborative document:

<https://notes.desy.de/pcZxHGIJQCGVr-vaDvITzg>



HedgeDoc

Using zoom

- interactive tools
- chat, hands, microphone

Rules

- Chat
- raise hands
- using document





# Welcome to workshop Archiving and Publishing at UFZ

---

Further documents:

Using RDM Guidelines

<https://rdm.pages.ufz.de/rdm-dissemination/rdm-guidelines/>

Visit HIFIS Workshops

<https://events.hifis.net/>

RDM & WKDV News (Mattermost Channel)

RDM Intranet Pages

HedgeDoc Document

# What are “research data”?

---

Samples: soil, water, tea bags  
Records  
Interviews  
Transcripts  
R codes  
Software  
Software licences  
Images  
GIS – spatial data  
Locations  
Building information  
Statistical data  
Energy consumption  
Databases  
Payment slips  
Permits  
Data protection agreements

Consent forms  
Parameters  
Conditions  
Log files  
Research diaries  
Lab journals  
E-Mails  
Results  
Simulations  
Input/output files  
Measurements  
Models  
Numbers  
Scenarios  
Tool settings  
Machinery, equipment  
Chemicals

Consumables  
Sequences  
Time series  
Remote sensing data  
Satellite data  
Indices  
Open access data  
Literature  
Micrographs  
Spectrographs  
Roughness measurements  
Presentations  
Loss protocols  
Destruction protocols  
Transfer protocols  
....



# Workshop agenda

---

1. Welcoming / getting to know each other / orientation
2. **Data publication in general**
3. Using UFZ- Data Management Portal (DMP) – Our Archive System
4. Using PANGAEA Data Publisher
5. Wrap-up and outlook



# Workshop agenda

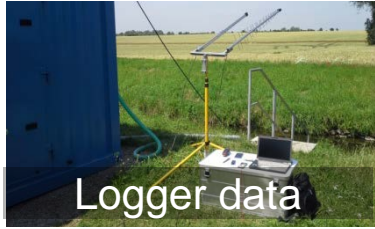
---

1. Welcoming / getting to know each other / orientation
2. Data publication in general
3. **Using UFZ- Data Management Portal (DMP) – Our Archive System**
4. Using PANGAEA Data Publisher
5. Wrap-up and outlook



# Tools - Publishing at UFZ

## Data Management Portal - Overview



### Datamanagementportal

UFZ Webapplikation

alle Datenprojekte

Holsteinstein | Lagerarbeiten | Umweltqualitätsdaten | Archivdaten | Beobachtungen

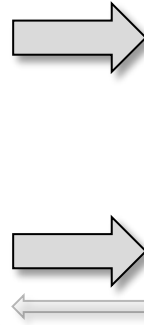
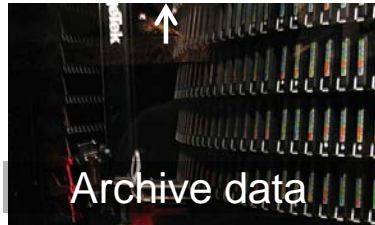
Willkommen im Datenmanagementportal (DMP)

Hier erfolgt die Verwaltung wissenschaftlicher Daten aus verschiedenen Bereichen:

- Archivdaten - Datenbanks Administration
- Probenkatalog - Abbildung von Probenahme- und Analysedatenworkflows
- Legenden - Automatische Übertragung von Messwerten von Geodaten im Feld
- Bewirtschaftungsdaten - Planung und Historie der Bewirtschaftung von Marschfeldern

Bei Fragen wenden Sie sich bitte an [info-datamanagement@ufz.de](mailto:info-datamanagement@ufz.de).

1.0.0.0 (aktuell) | 1.0.0.0 (aktuell) | 1.0.0.0 (aktuell) | 1.0.0.0 (aktuell)



### DataInvestigationPortal

Probenorten | Lagerarbeiten | Umweltqualitätsdaten | Beobachtungen

LCMS_LTQOnline_Selene_Kuechler_Holtenauer_LVSPE_2014_EDreg_target_T323H047	Details	Download
LCMS_Quantitative_Martin_Krauss_Holtenauer_LVSPE_2015_rennerget_EDSipon_T323H047_160211	Details	Download
LCMS_LTQOnline_Lisa-Marie_Krauskopf_Holtenauer_surfactants_standards_T323H047_160121	Details	Download
Wassergütemessstation Holtenauer (Conductivity sensor in LV/SPE device (MC Nierthagen))	Details	Download
Holtenauer (378-H0518)	Details	Download
Holtenauer (Kuck_340)	Details	Download
Wassergütemessstation Holtenauer (Conductivity sensor in LV/SPE device (MC Nierthagen))	Details	Download

**DOWNLOAD**

### Interfaces

SQL

R

FZ



# Tools

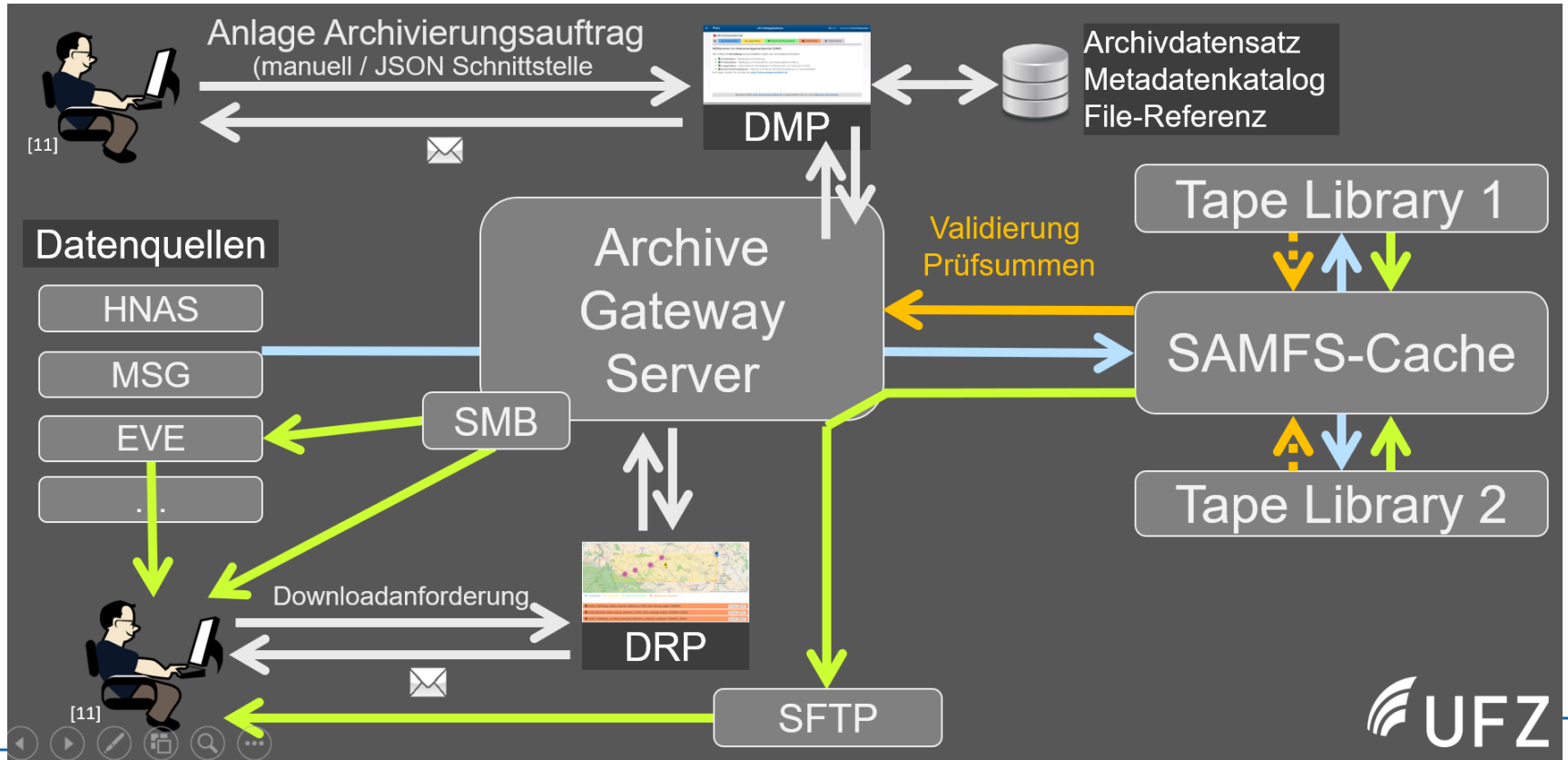
## Archive data - performance characteristics

---

- File archiving for an **unlimited** period
- Support for large amounts of data (terabyte range)
- Description of the data using **Dublin Core**
- Mandatory filing of **license** and usage **conditions**
- Ensuring **data integrity** through checksums **Protection against data loss** (locally separate archives)
- **Access** via file explorer or data research portal
- **Provision** of the data to a configurable group of users (including external cooperation partners)
- **DOI Processing**

# Tools

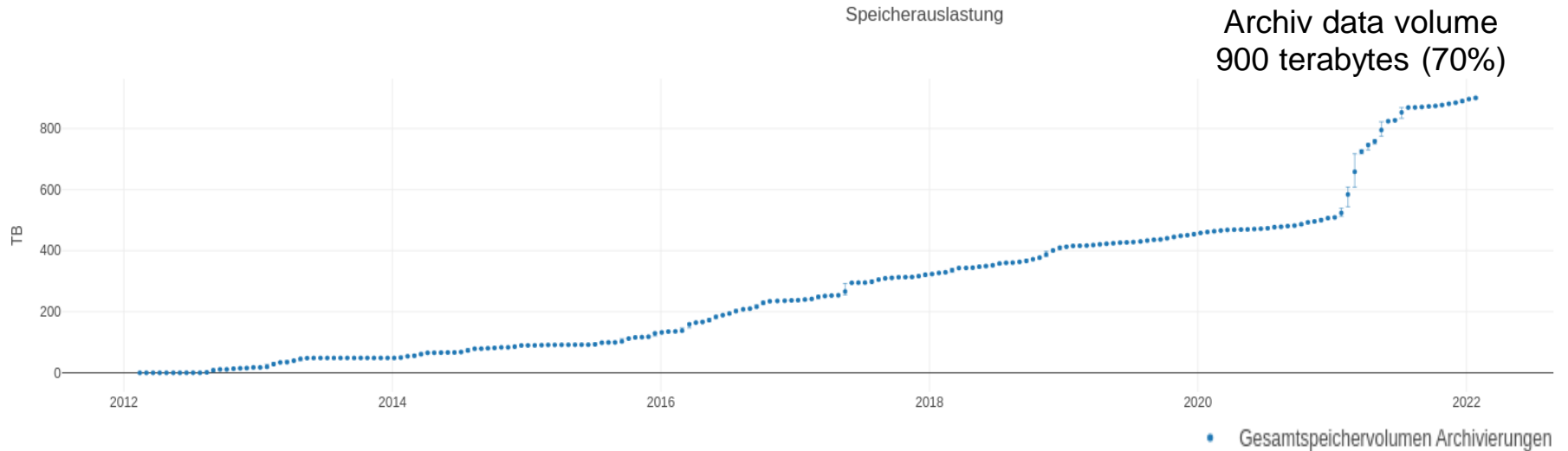
## Archivdaten – Archivierungsprozess / Datenretrieval





# Tools

## Data Management Portal – Archive System



- free storage memory in the archive system: 400 terabytes (30%)
- hardware renewal in 2022
- increase to around 720 terabytes free memory over the year

# Tools

## Data Management Portal – Archive System

---

<https://webapp.intranet.ufz.de/wkdv/grafana/d/x2Gv60u7zUFZ/dmp-archive-all-time-stats?orgId=2&refresh=1d&from=now-10y&to=now>

# Tools

## Data Management Portal - Access

- UFZ -> Internal Area -> web application
- **Data Management Portal**

The screenshot shows the UFZ Web Applications interface. At the top, there is a dark blue header with the UFZ logo on the left, the text "UFZ Web Applications" in the center, and "DE | EN" on the right. Below the header, there is a navigation bar with a dropdown menu set to "all data projects" and five buttons: "Sample data" (blue), "Logger data" (yellow), "Field management data" (green), "Archive data" (orange), and "Master data" (grey). The main content area starts with the heading "Welcome to the Data Management Portal (DMP)". Below this, a paragraph states: "With this application you are able to **manage** scientific data from different areas:". This is followed by a bulleted list of four items, each with a green checkmark icon: "Archive data - File based archiving", "Sample data - Sampling and analysis data workflow", "Logger data - Automated transmission of measured values from sensors in the field", and "Field management data - Planning and history of agement on experimental fields". At the bottom of the main content area, there is a line of text: "In case of any question please contact [wkdv-datamanagement@ufz.de](mailto:wkdv-datamanagement@ufz.de)".

UFZ Web Applications DE | EN

all data projects

Sample data Logger data Field management data Archive data Master data

### Welcome to the Data Management Portal (DMP)

With this application you are able to **manage** scientific data from different areas:

- **Archive data** - File based archiving
- **Sample data** - Sampling and analysis data workflow
- **Logger data** - Automated transmission of measured values from sensors in the field
- **Field management data** - Planning and history of agement on experimental fields

In case of any question please contact [wkdv-datamanagement@ufz.de](mailto:wkdv-datamanagement@ufz.de).

## Data Management Portal – Archive Data

The image shows two screenshots of the UFZ Web Applications interface. The top screenshot shows the main navigation bar with the UFZ logo, 'UFZ Web Applications', and language options 'DE | EN'. Below the navigation bar, there are several buttons: 'all data projects', 'Sample data', 'Logger data', and 'Field data'. A red box highlights the 'Archive data' button, with a black arrow pointing down to the second screenshot.

The second screenshot shows the 'Archive data' interface for the project 'Statischer Dauerdüngerversuch / Static Fertilization Experiment Bad Lauchstädt'. The interface includes a search bar, navigation tabs for 'Probanddaten', 'Loggerdaten', 'Bewirtschaftungsdaten', 'Archivdaten', and 'Stammdaten'. The main content area is titled 'Archivdatensätze und Archivierungsaufträge des ausgewählten Datenprojekts' and displays system status, draft records, pending archive orders, and a list of successfully completed archive operations.

**Systemauslastung:**

- 1 Archivierungsaufträge mit einem Volumen von 8,310 GB werden gerade archiviert
- keine Archivierungsaufträge warten auf die Verarbeitung

**Entwürfe (0)**

Anzeige: alles | Sortierung: Datensatzbezeichnung | Seite 1 von 1

*Es sind keine Datensätze vorhanden, die den Kriterien entsprechen.*

**Archivierungsaufträge (0)**

Anzeige: alles | Sortierung: Archivierungsstatus | Seite 1 von 1

*Es sind keine Datensätze vorhanden, die den Kriterien entsprechen.*

**Erfolgreich abgeschlossene Archivierungen (34)**

Anzeige: alles | Sortierung: Anlagendatum | Seite 1 von 4

StaticFertilizationExperiment_BadLauchstaedt_2001-2014_SH7_SH8	[Icons]
StaticFertilizationExperiment_BadLauchstaedt_2001-2014_SH2_SH3_SH6_SH7	[Icons]
StaticFertilizationExperiment_BadLauchstaedt_1991-2000_SH7_SH8	[Icons]

# Tools

## Data Management Portal – Archive Data

[HICAM] climate projection; meteorological data; pr met\_003, EU domain quality assured data


**Assignment to data project:**


POF3 - T53 - From models to predictions - HICAM

**Data retrieval:**

[Request data for download via Data Investigation Portal](#)  
(recommended for large data)

**Access data via file explorer**  
(only within the UFZ network / recommended for small data)

 Y:\Archive\POF3 - T53 - From models to predictions - HICAM - 609\\_HICAM\\_ climate projection\\_ meteorological data\\_ pr met\_003\\_ EU domain - 9332\

 smb://archive-gateway.intranet.ufz.de/archive/POF3 - T53 - From models to predictions - HICAM - 609/\_HICAM\\_ climate projection\\_ meteorological data\\_ pr met\_003\\_ EU domain - 9332/

**Description of data:**

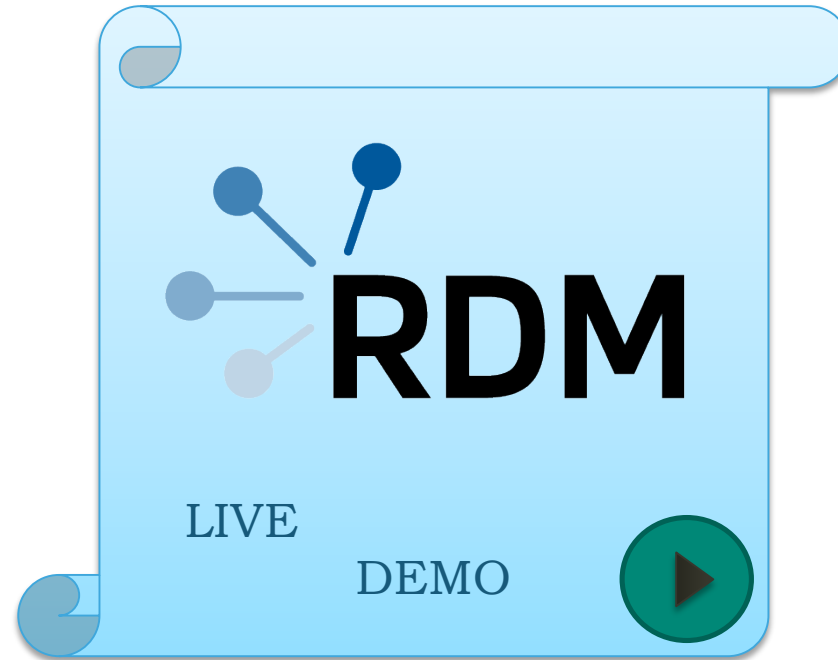
**Dublin Core**

<b>URI:</b>	http://www.ufz.de/record/dmp/archive/9332
<b>Title:</b>	[HICAM] climate projection; meteorological data; pr met_003, EU domain
<b>Subject:</b>	climate projections, meteorological forcing data; pr met_003
<b>Description:</b>	Back up of /data/hicam/data/processed/meteo/europe/climproj/euro-cordex/88realizations_fromGERICS/prAdjust_EUR-11_CNRM-CERFACS-CNRM-CM5_rcp26_r1i1p1_CNRM-ALADIN63_v2-GERICS-ISIMIP3BASD-UFZ-3km-1971-2000_day_19710101-20991231.nc



# Lets take a look: Data Management Portal

## *Archive data – Live Demo*



# Tools

## Archive Data – automated archiving process

---

Download **import file**. (See specification for structure of import file)

Copy “**.import**” file to \\archive-gateway.intranet.ufz.de\interface\USERNAME

Linux and Mac: smb://archive-gateway/interface/USERNAME

EVE cluster: /archive-interface/USERNAME/USERNAME/

System changes file ending in “**.validating**” -> wait 10 minutes. (“first come, first served”).

System create new **archiving order**.

Should the import file be formally or substantitvely invalid, file ending renamed into “**.invalidformat**” or “**.invaliddata**” and the import process aborted.

If archiving order was successful -> file ending “**.processing**”.

# Tools

## Archive Data – automated archiving process

---

Check web interface (Datamanagementportal)

Delete archive within 1 hour if necessary —> file ending "**.deletedbyuser**"

If archiving failed -> file ending "**.error**" (and information by email)

If archiving was successful, "**.success**" (and email note)

<https://www.intranet.ufz.de/index.php?de=44993>

# Tools

## Data Investigation Portal - Example

ASD-Spektren, Licor 6400-Daten, Dissertation Benjamin Dechant 2013



**Projekt** > AISA-Hyperspectral

**Beschreibung** Dissertation Benjamin Dechant  
ASD-Spektren und Gasdurchflußmessungen diverser baumarten,  
Arboretum Großpösna Herbst 2013

**Größe** 1,022,8 MB

➤ [Daten zum Download anfordern](#)

➤ [Kontakt](#)

▼ [Metadaten Kataloge](#)

▼ [Dublin Core](#)

**URI** <http://www.ufz.de/record/dmp/archive/7517>

**Thema** ASD-Spektren, Licor 6400 Messdaten

**Geltungsbereich** Großpösna 2013



---

**Time for a  15 min break!**



# Workshop agenda

---

1. Welcoming / getting to know each other / orientation
2. Data publication in general
3. Using UFZ- Data Management Portal (DMP) – Our Archive System
4. **Using PANGAEA Data Publisher**
5. Wrap-up and outlook



# Workshop agenda

---

1. Welcoming / getting to know each other / orientation
2. Data publication in general
3. Using UFZ- Data Management Portal (DMP) – Our Archive System
4. Using PANGAEA Data Publisher
5. **Wrap-up and outlook**



# Conclusion

---

Analyze the type and volume of your data!

How you can structure your data? Which Metadata is needed?

Do you need a persistence Identifier (PID, DOI)?

Are there requirements to long-term availability of your data?

Keep FAIR in your mind!

**How FAIR we are at UFZ?**

**How FAIR you are?**



# How to connect and engage?

- Motivate your colleagues
- Share your own best practices
- set a example



[ Graphic: [Scriberia](#) under [CC-BY](#) licence ]



---

## Thank you for your interest!

This course partially uses the concept and material of the “Train-the-Trainer” concept of the FDMentor project.

<http://doi.org/10.5281/zenodo.4322849>



Material was also used from GFBio Roadshow 2020 by Tina Astor and from Research Data Management Workshop “Make your data count!” 2017 by Birgitta König-Ries and Roman Gerlach.

Presentation used from Janine Felden, Frank Oliver Glöckner & PANGAEA Team about Pangaea from RDA 2022 workshop