

# **Forschungsdatenmanagement aus der Sicht eines INF-Projektbearbeiters**

**Campustreffen Forschungsdaten  
23. Juli 2020**

Harald Lanig

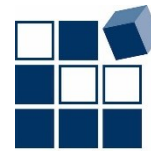
Diplom-Chemiker, Modellierung und Simulation von Proteinstrukturen

Wissenschaftlicher Mitarbeiter am

**Zentralinstitut für Scientific Computing**

Verantwortlich für das Forschungsdatenmanagement in

**SFB 814 „Additive Manufacturing“**



**SFB 814**  
ADDITIVE FERTIGUNG

**SFB 1411 „Design of Particulate Products“**



Öffentliche (BMBF, ERC, DFG) und private Geldgeber verlangen zunehmend Konzepte zur **nachhaltigen Wiederverwendung** von Forschungsergebnissen

Der Begriff **Wiederverwendung** (-verwendbarkeit) bedingt die

Sicherung der Datenintegrität (Informationssicherheit durch **data provenance**)

Annotieren der Forschungsdaten mit **Metainformationen**

- Beschreiben den Kontext der Datenerhebung
- Versuchs- oder Simulationsbedingungen
- Analysen, Auswertungen, begleitende Dokumentation
- Definition von Prozessketten

**Speicherung** der Daten

- Während der Projektlaufzeit (dabei kontinuierliche Nutzung)
- Langfristig (unveränderlich, mit DOI, Supp Inf zu Publikationen)

„Die DFG ist [...] bestrebt, durch ihre Förderung auch zur **Sicherung, Aufbewahrung und Nachnutzbarkeit von Forschungsdaten** beizutragen.“

DFG-Vordruck 54.01 – 02/20 (Leitfaden Antragstellung Projektanträge)

## Übergeordnete Leitlinien der DFG

### 1. Projektplanung und Antragstellung

- Welche Daten entstehen, werden erzeugt oder ausgewertet?
- Informationen zu Datentypen, disziplinspezifischen Standards und geeigneter Repositorien. Berücksichtigung fachspezifischer Konzepte

### 2. Bereitstellung

- Ermöglichung der **sinnvollen Nach- und Weiternutzung**, sofern die Rechte Dritter (Datenschutz, Urheberrecht) dies erlauben

### 3. Langfristige Sicherung

- **Archivierung in der eigenen Einrichtung** oder einer fachlich einschlägigen, überregionalen Infrastruktur **für mindestens 10 Jahre**

[www.dfg.de/antragstellung/forschungsdaten](http://www.dfg.de/antragstellung/forschungsdaten)

# SFB-Antragsmuster

## 1.4.3 Forschungsdaten/Wissensmanagement

Die Verbesserung des Umgangs mit Forschungsdaten hat sowohl bei nationalen und internationalen Forschungsorganisationen als auch in der Wissenschaft eine hohe Bedeutung. Die DFG ist daher bestrebt, durch ihre Förderung auch zur Sicherung, Aufbewahrung und nachhaltigen Verfügbarkeit der Forschungsdaten beizutragen (vgl. DFG-Leitlinien zum Umgang mit FD).

Bitte führen Sie aus, welche Forschungsdaten im SFB entstehen, erzeugt oder ausgewertet werden. Legen Sie dabei bitte fachspezifisch angemessene **Konzepte und Überlegungen für die Qualitätssicherung, für den Umgang mit und die langfristige Sicherung der Forschungsdaten** zugrunde. Die einschlägigen Erläuterungen müssen Informationen zu Datentypen, falls vorhanden zu **disziplinspezifischen Standards** und zur **Wahl geeigneter Repositorien** enthalten, sofern diese für ein bestimmtes Fachgebiet oder bestimmte Datentypen vorhanden sind. Zusätzlich werden Angaben zu ggf. **betroffenen Rechten Dritter** sowie erste Planungen zum Zeitpunkt der Datenveröffentlichung erbeten. Stellen Sie bitte auch dar, durch welche **Einrichtungen der beteiligten Institutionen** (Datenkurator, Rechenzentrum, Bibliothek, etc.) welche Form von Unterstützung beim Daten- und Informations-/ Wissensmanagement geleistet werden soll.

**--- FDM als Teil von Z oder besser ein eigenes Teilprojekt INF? ---**

# Key challenges transforming data into knowledge

- **Digitalization** of laboratory notebooks and simulation protocols
- **Seamless integration** of experimental and simulation data
- **Documentation** of the whole history of measured and simulated data
- **(Long-term) archiving** and accessibility of research data

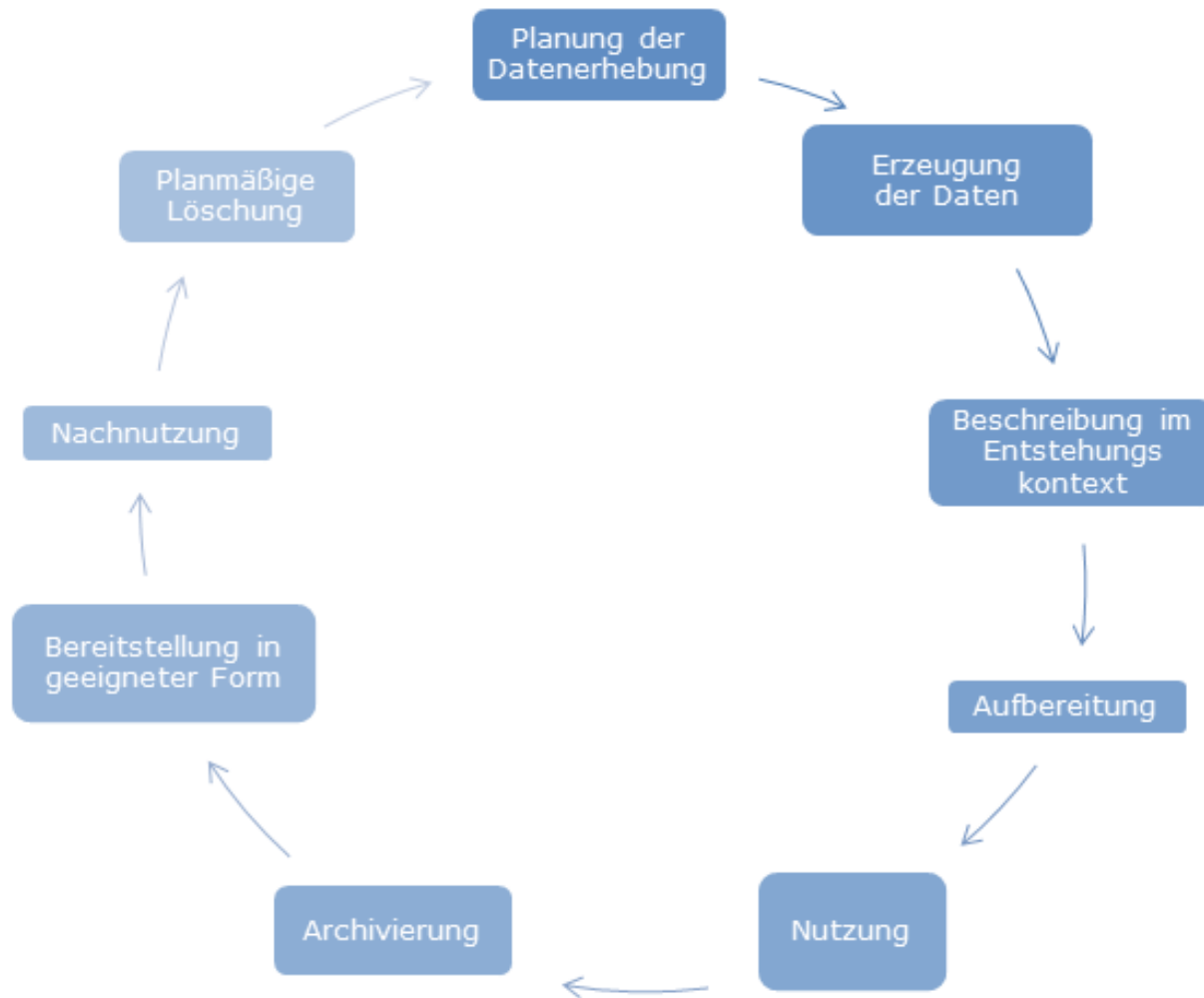
These goals can be achieved by **establishing a research environment** which enables **effective information and data transfer** within and between projects.

This requires a **systematical data management** within the CRC **according to FAIR principles**. They include

the capacity of computational systems to

**find, access, interoperate, and reuse data**

with none or minimal human intervention.



Wie können möglichst viele dieser Aspekte abgedeckt werden?

[www.fdm-bayern.org/forschungsdaten-lebenszyklus/](http://www.fdm-bayern.org/forschungsdaten-lebenszyklus/)

The use of an **Electronic Lab Notebook** and **Lab Information Management** system allows the researcher to document the work, describe materials and methods by annotating everything with metadata, and to collect raw, analyzed, or processed data.

The goal of such a system is

**supporting the workflow of research data  
from generation to publication.**

In every step, data should be accessible, editable, shareable.  
This allows the efficient (re-)use of raw or processed data.

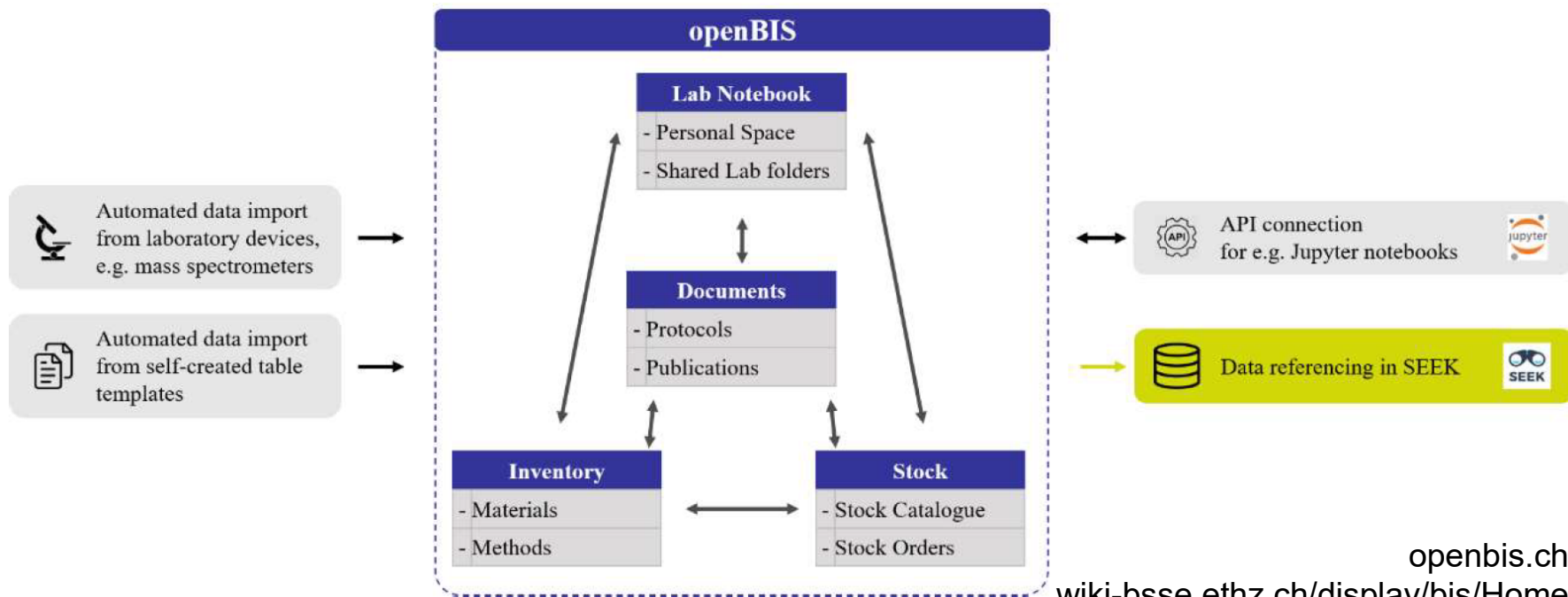
Using an ELN system ensures that even when the researcher is leaving the lab, the **metadata** and the underlying **raw data** is **still present**.



# openBIS

## a configurable ELN-LIM system

- Open source system, developed by Scientific IT Services (SIS) at ETH
- Store and document (annotate) your research data
- Basic areas **Lab Notebook**, **Documents**, **Inventory**, and **Stock** are organized in **Spaces** (= folders), which are **hierarchically structured**
- Relationship between objects is defined by parent-child-inheritance
- Configurable access and rights management





### INVENTORY MANAGEMENT

Whatever your research field, keep track of all materials and samples used in your lab. Create experimental or computational procedures to share with all lab members.

#### INVENTORY MANAGEMENT



### BIGDATALINK

Use openBIS as metadata repository and use our git-like tool to link huge datasets across different locations.

#### BIG-DATA LINKAGE



### ACCESS AND RIGHTS MANAGEMENT

Full control of user profile and access privileges.

#### RIGHTS MANAGEMENT



### LAB NOTEBOOK

Describe your wet-lab or computational experiments and link to the materials, samples and protocols stored in the inventory.

#### LAB NOTEBOOK



### INTEGRATIONS FOR DATA ANALYSIS

Use Jupyter notebook to analyse your data and store your notebooks in openBIS. Use the APIs to integrate openBIS in your workflow manager to extract data for large-scale analysis on a computer cluster and upload results back to openBIS.

#### DATA ANALYSIS



### AUDIT TRAIL

Keep track of all modifications entered in the database.

#### TRACK HISTORY



### DATA MANAGEMENT

Make your data FAIR. Store all data connected to your experiments, of any size and format, via the web interface or using background upload scripts, depending on the data size.

#### DATA MANAGEMENT



### IMPORT/EXPORT

Excel-compatible import and export formats for all tables. Export of lab notebook to text and word files.

#### IMPORT / EXPORT

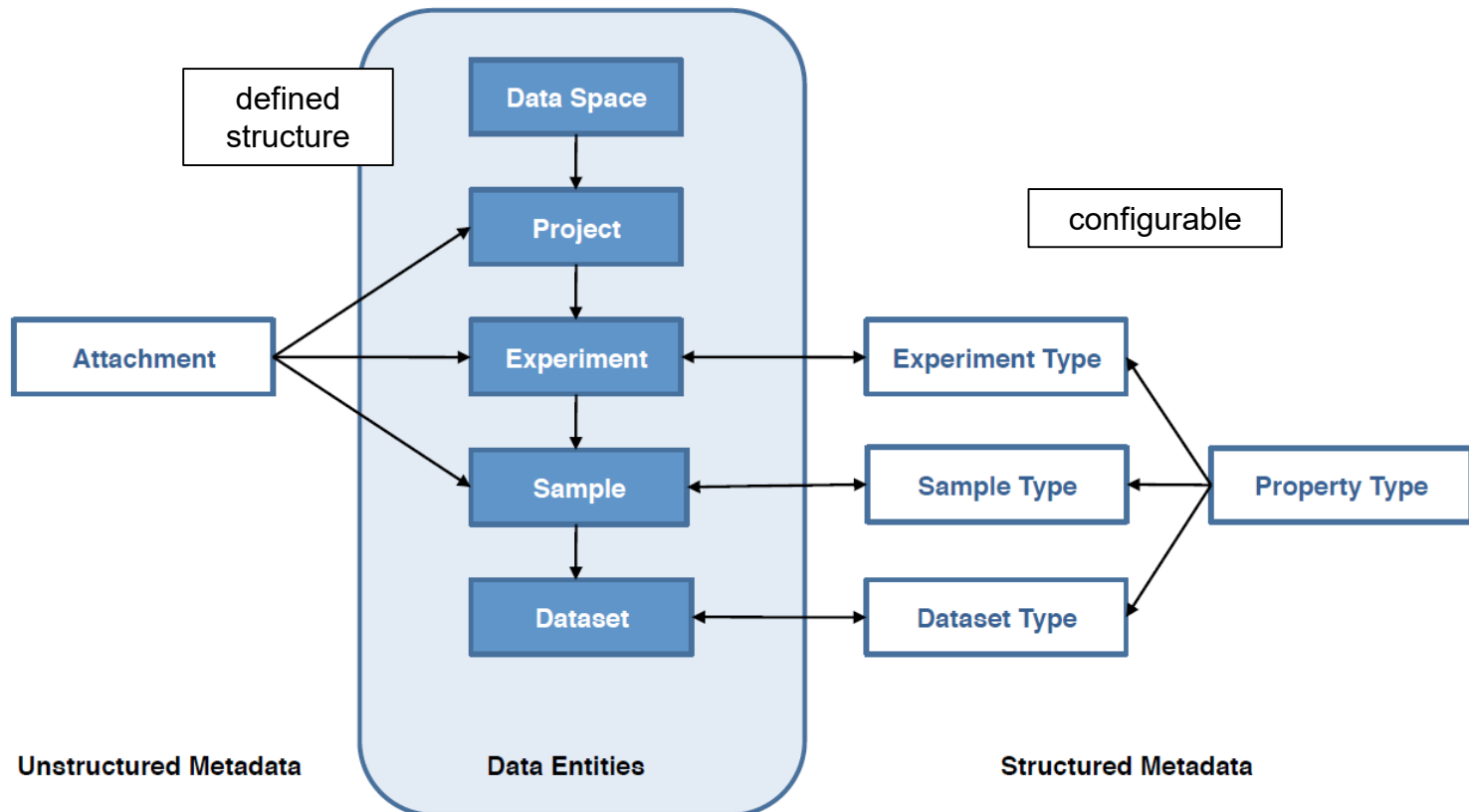


### MODULARITY

Write your own plugins to extend existing functionalities or add new ones.

#### MODULARITY

- Default **Data Space**, e.g. for Materials and Methods, or Lab Space
- Concept of **samples (=objects)** linked to corresponding **data**, grouped into **experiments (=collections)**
- Annotation by **structured custom properties** and individual **attachments**



## Steps to establish an ELN within a CRC

- Setup a local ELN server (also with limited storage available)
- Allocate a few PIs with test projects
- Define templates for objects, procedures, methods, experimental steps
  - already existing data(bases)
  - lab protocols (synthesis/measurement/analysis)
  - data workflow description  
acquisition → processing → analysis
- Test the ELN in parallel (yes, this is extra work)

## The use of an ELN will provide

- an environment for **combination of experimental and simulation data** and **data exchange** in ongoing projects and collaborations
- an infrastructure for simple **data accessibility** (and archiving)
- an easy way of addressing **FAIR data** principles

**Vielen Dank für die Aufmerksamkeit!**

**Campustreffen Forschungsdaten  
23. Juli 2020**

Harald Lanig