

# NFDI4Health Local Data Hub

Extended metadata feature of SEEK in the Local Data Hub of NFDI4Health

Frank Meineke

Martin Golebiewski

Xiaoming Hu

FAIRDOM User Meeting

5.6.2024



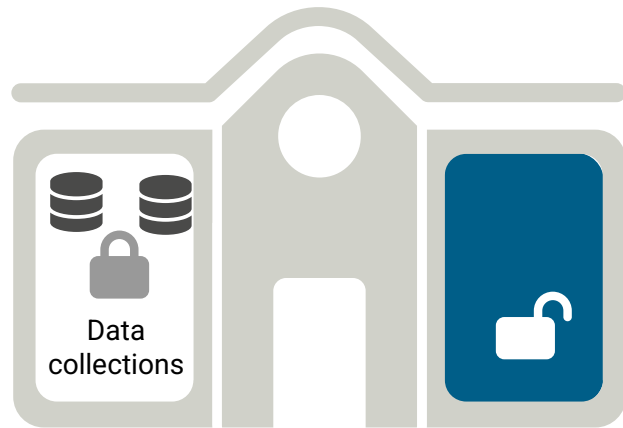
UNIVERSITÄT  
LEIPZIG

Heidelberg Institute for  
Theoretical Studies



# The NFDI4Health consortium

- Research institutes involved in data collection and analysis of personal health data
- Infrastructures / experts that develop data standards and methods
- Institutions with focus on community involvement
- 47 renowned institutions as participants
- Develops central and decentral Services for



Research data centre

**Synonym: Data Holding Organisation (DHO)**



# NFDI4Health Services - central and decentral components

German central  
Health Study Hub

Findability

## Search & Exploration \*

- Search from variable definitions to study protocols and scientific publications at multiple levels of granularity
- Unify and aggregate multiple data repositories to enable search across institutional boundaries

## Data publication

- Publication of sensitive research data without compromising data protection
- Assignment of persistent identifier (digital object identifier (DOI)) possible

\* No sensitive personal level data included

Local Data Hub

Findability + Reuse

## Business card of the organisation (Yellow Pages)

- Showcase an institution's ongoing and completed work (Projects/Studies/Datasets/...)
- Provides control over the information displayed through fine-grained access rights management

## Sensitive Research Data Management

- Creating links between metadata and data in secure data silos
- Interconnects with local data analytics components like DataSHIELD/ Personal Health Train

## Transfer Unit

- Facilitate local use and reuse of data through granular management of access rights

Central Access  
Portal

Accessibility + Reuse

## Data Request

- Request access to sensitive data of other Data Holding Organisations (not within home institution)
- Apply for multiple datasets with one request

## Communication & Support

- Unified communication channel between all involved parties (Data Holding Organisations and data requesting parties)

## Transparency Portal

- View & explore granted data use requests

# Interaction from the researcher's point of view

German central  
Health Study Hub

Local Data Hub

Central Access  
Portal

1

## Find relevant studies

- Multifaceted search across all registered studies
- Follow link to Local Data Hub of original source

2

## Find out more!

- Discover the environment (projects, people, etc.) and opportunities for contact and data sharing
- Reach public or data directly or request confidential data

3

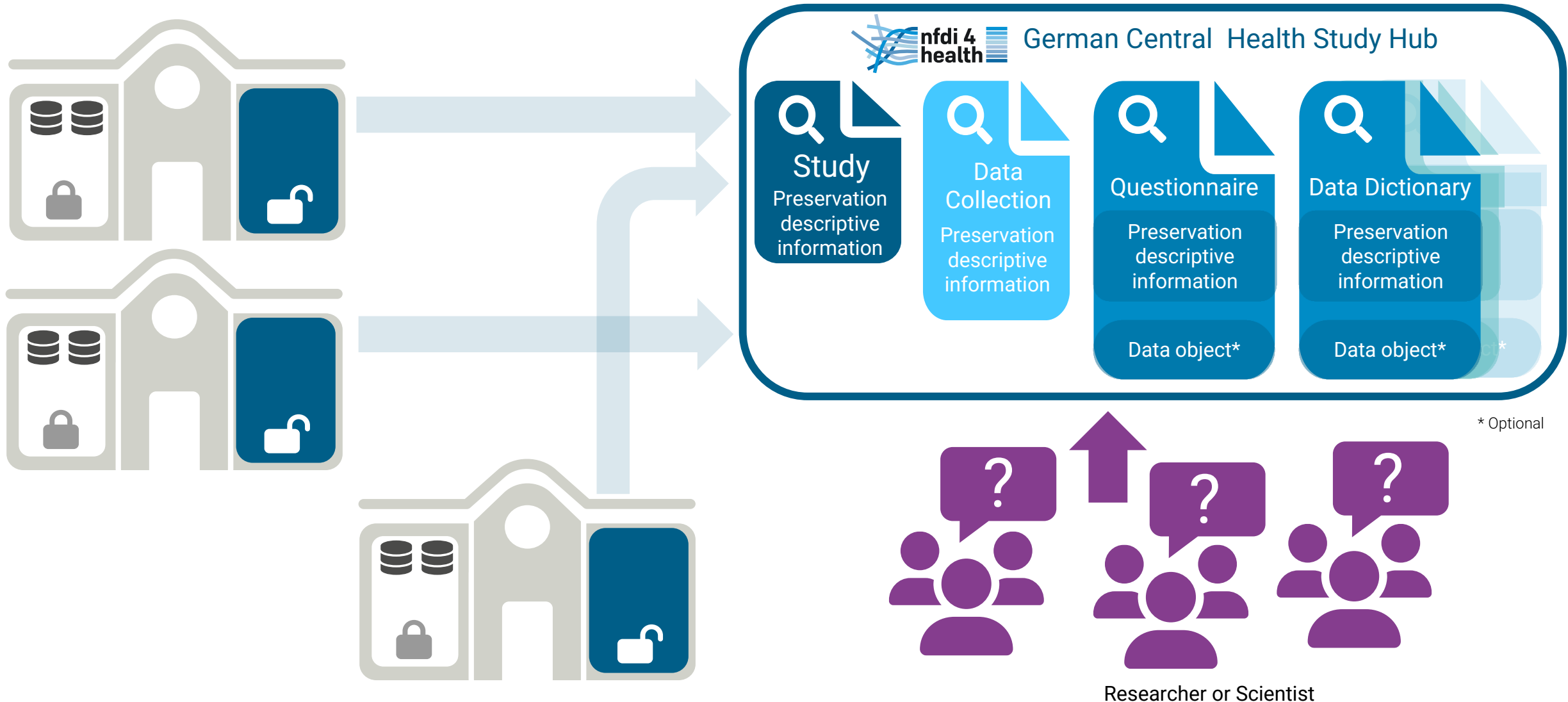
## Get Data Access

- Request data for centralized or distributed analysis from multiple centers
- e.g. negotiate DataSHIELD or Personal Health Train access



Researcher

# Findability German Central Health Study Hub – Data provider

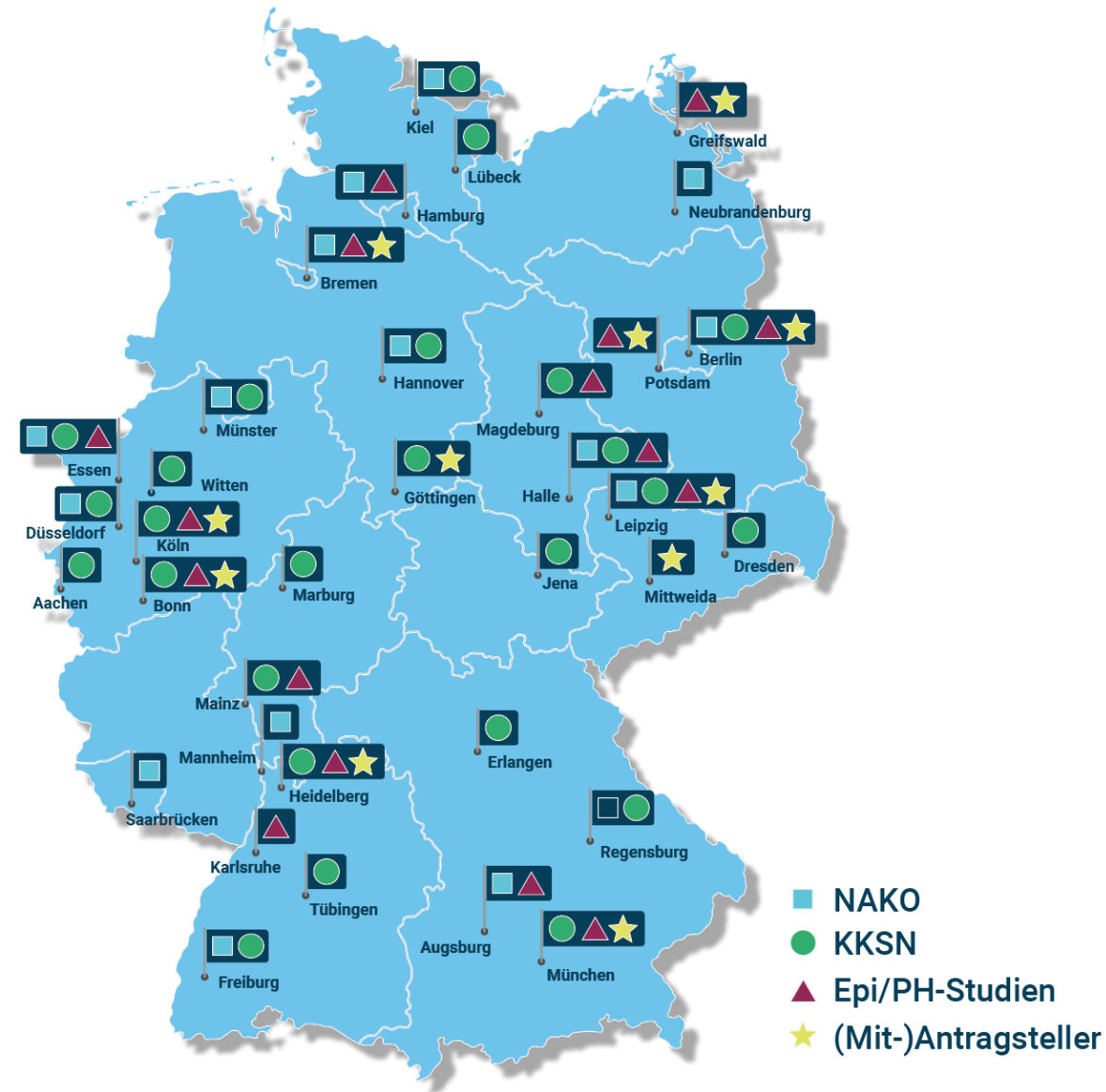


# Rollout of Local Data Hubs

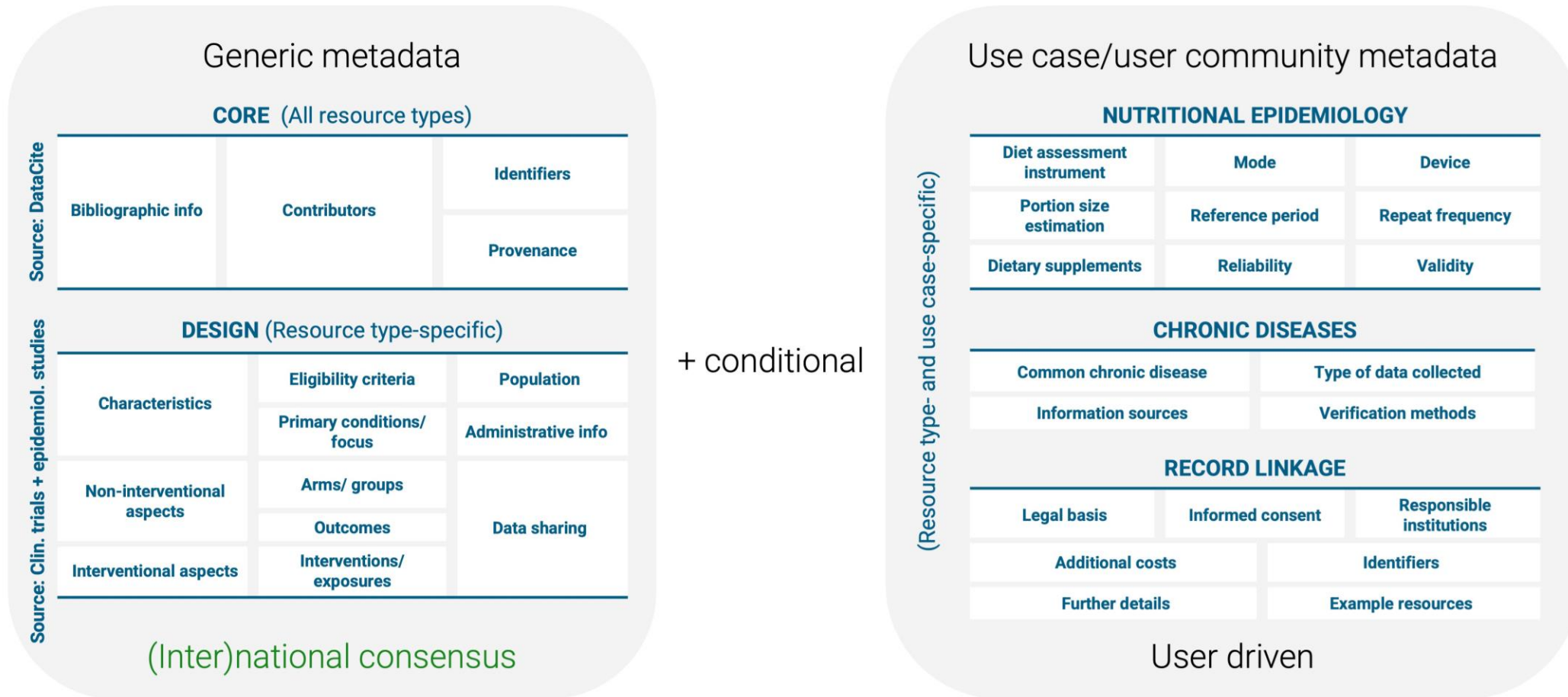
- Epidemiological and public health studies ▲
- 26 local studies with > 400.000 participants
  - Incl. German National Cohort (NAKO) ■
- 24 university study centres ●

Each Data Holding Organisation should be linked to an LDH/SEEK

- Locally installed (self cared)
- Hosted (IMISE will host about 10 LDHs)
- Shared (with other DHO, like health-atlas.de)



# NFDI4Health Metadata Schema / Modul Overview



# NFDI4Health Metadata Schema / Specification

- Core: 82 lines/ Design 112 lines of Excel
- Coming up: ArtDecor: <https://art-decor.org/ad/#/nfdhtfcov19-/datasets/dataset>

No 3_2	No 3_3	Item	Data Type	Cardinality	Display information? (Data capturing)	Display information? (Data visualization)	Allowed Values	Data element heading    Short name to display (display_name)	Data element description (description)
1.11	1.11	Resource.idsAlternative	BackboneElement	0..*	yes	yes	-	Alternative identifiers	Group of items providing information about identifiers (IDs) assigned to the given [RESOURCE] by another registering systems, e.g. a registry of clinical trials or a data repository.
1.11.1	1.11.1	Resource.idsAlternative.scheme	CodeableConcept	1..1	yes	yes	DRKS NCT (ClinicalTrials.gov) ISRCTN EudraCT EUDAMED UTN KonsortSWD MDM Portal Other	Type of the registry	Type/name of the system where the given [RESOURCE] is already registered.
1.11.2	1.11.2	Resource.idsAlternative.identifier	string	1..1	yes	yes	-	Identifier	Identifier (ID) assigned to the given [RESOURCE] by another registering system, e.g. by a registry of clinical trials or a data repository.

# What is Extended Metadata.. (formerly: custom metadata)

- Resources in SEEK have only very few and basic fixed attributes

Programme: *Independent Projects*

FAIRDOM PALs: *No PALs for this Project*

SEEK ID: <https://fairdomhub.org/projects/282>

Project created: 17th Feb 2022

Public web page: *Not specified*

Organisms: *No Organisms specified*

- Adding own attributes would need direct development
- Custom Metadata allows FAIRDOM-SEEK entities to have their metadata extended with key/value pairs, controlled by metadata types that can be validated, and set as required or optional (introduced in 1.11.0 / 2021)
- see [MIAPPE, Plant Data Integration and Findability](#)

## Investigation

Extended metadata ▾

**Id \***

**Submission date**

**License**

**Miappe version \***

## Study

Extended metadata ▾

**Id \***

**Study start date \***

**Study end date**

**Contact institution \***

**Geographic location country \***

**Experimental site name \***

**Latitude**

**Longitude**

**Altitude**

# Extended Metadata /

Until 2023

- Seek offered only a flat list of attributes – not helpful for our use case
- We needed an CRF (Case Report Form) like survey
- We started developing a completely new resource type
- but **abandoned** that proprietary approach in favour of improvement of EXMD feature for whole community

# Extended Metadata /

Until 2023

- Seek offered only a flat list of attributes – not helpful for our use case
- We needed an CRF (Case Report Form) like survey
- We started developing a completely new resource type
- but **abandoned** that proprietary approach in favour of improvement of EXMD feature for whole community

From 2023/24 we implemented possibility

- Adding multi-select from code lists
- Grouping of attributes
- Nesting of groups
- Repeating of groups (dynamic, with add/remove)

Eligible gender

Male  Female |

Male

Female

Diverse

Not applicable

Alternative identifiers ▲

Type of the registry(\*)

EudraCT

Identifier(\*)

2013-004882-15

Remove

Type of the registry(\*)

DRKS

Identifier(\*)

DRKS00006240

Remove

# NFDI4Health Metadata Schema / Seed File

- Seeds.rb is generated from excel by a proprietary python script
  - 546 lines of ruby code (e.g. MIAPPE has only 40 lines...)
  - Threefold, as identifier have to technically resource specific (here: project, investigation, study)
  - [https://github.com/nfdi4health/ldh/blob/main/db/seeds/018\\_MDS\\_3\\_3\\_Project\\_gen.seeds.rb](https://github.com/nfdi4health/ldh/blob/main/db/seeds/018_MDS_3_3_Project_gen.seeds.rb)
- Technically challenges
  - This is imported (“seeded”) on start or when manually triggered
  - It is integrated into our deployment docker image (docker build)
  - updates require re-deployment (docker pull new release)
  - process might be optimizable (e.g. external mount of seeds, outside container)
  - no explicit versioning on SEEK side, no migration support

**Resource classification(\*)** ▾

- **Type of the resource(\*)** : Study

**Title(s)/name(s) of the Project** ▾

- **Title/name(\*)** : A multicentre randomized controlled trial evaluating the rate of sustained treatment in non-cirrhotic HBeAg-negative chronic Hepatitis B patients with long-term viraemia
- **Language of the title/name(\*)** : English

**Acronym(s) of the Project** ▾

- **Acronym(\*)** : STOP-NUC
- **Language of the acronym(\*)** : English

**Description(s) of the Project** ▾

- **Description(\*)** : Nucleos(t)ide analogues (NUCs) are the standard and mostly lifelong treatment for chronic hepatitis B (CHB). Sustained loss of HBsAg (loss of HBsAg) is rarely achieved. Discontinuation of NUC treatment may lead to relapse based on small or non-randomized clinical trials. The STOP-NUC trial was designed with treatment interruption approach. In this multicenter, randomized controlled trial, 166 long-term NUC treatment, with HBV DNA < 172 IU/mL (1,000 copies/mL) for at least 4 years treatment (Arm B) for a 96-week observation period. In total, 158 patients were available sustained HBsAg loss up to week 96.
- **Language of the description(\*)** : English

**Keyword(s) describing the Project** ▾

- **Keyword(\*)** : chronic HBeAg-negative hepatitis B

**Language(s) of the project** : English

**Web page of the project** : <https://www.clinicaltrialsregister.eu/ctr-search/trial/2013-004882-15/1>

**Contributor(s) of the Project** ▾

- **Is it a personal or organisational contribution?(\*)** : Personal
- **Details about the contributing organisation(s)/institution(s)/group(s)** ▾
  - **Contributor type(\*)** : *Not specified*
  - **Funding identifier(s)** : *Not specified*
  - **Name of the organisation/institution/group(\*)** : *Not specified*
- **Details about the contributing person(s)** ▾
  - **Contributor type(\*)** : Principal investigator
  - **Given name(\*)** : Florian
  - **Family name(s) (\*)** : van Bömmel
  - **Digital identifier(s)** ▾
    - **Identifier(\*)** : *Not specified*
    - **Scheme(\*)** : *Not specified*
- **Email address** : Florian.vanBoemmel@medizin.uni-leipzig.de
- **Phone number** : +4934197 12200
- **Organisation(s) associated with the contributor** ▾
  - **Name(\*)** : Universität Leipzig
  - **Address** : Liebigstraße 20, 04103 Leipzig
  - **Web page** : *Not specified*
  - **Digital identifier(s)** ▾
    - **Identifier(\*)** : *Not specified*
    - **Scheme(\*)** : *Not specified*

**Alternative identifiers** ▾

- **Type of the registry(\*)** : EudraCT
- **Identifier(\*)** : 2013-004882-15
- **Type of the registry(\*)** : DRKS
- **Identifier(\*)** : DRKS00006240

**Includes chronic diseases?** : true

**Characteristics of the Project** ▾

- **Is it an interventional or non-interventional Project?** : Interventional
  - **Specification of the type of the Project** ▾
    - **Interventional Project type** : Unknown
    - **Non-interventional Project type** : []
  - **Primary health condition(s) or disease(s) considered in the Project** ▾
    - **Primary health condition or disease name(\*)** : *Not specified*
    - **Terminology/classification(\*)** : *Not specified*
    - **Code of the health condition or disease** : *Not specified*
  - **Groups of diseases or conditions(\*)** ▾
    - **Which groups of diseases or conditions were the data collected on?** : Certain infectious or parasitic diseases (I)
    - **On which diseases or conditions were the data collected?** : []
  - **Collects mortality data?** : *Not specified*
  - **Administrative information about the Project** ▾
    - **Status of the ethics committee approval** : Request for approval submitted, approval granted
    - **Overall Project status** : Completed: Recruitment, data collection, and data quality management completed normally
    - **Participants enrolled by invitation?** : *Not specified*
    - **Start date** : *Not specified*
    - **End date** : *Not specified*
  - **Mono- or multicentric?** : Multicentric
  - **Number of centers** : 24
  - **One or more data providers?** : *Not specified*
  - **Number of data providers** : *Not specified*
  - **Primary subject(\*)** : Person
  - **Eligibility criteria for Project participants** ▾
    - **Eligibility criteria: Minimum age** ▾
      - **Minimum eligible age(\*)** : 18
      - **Unit of age(\*)** : Years
    - **Eligibility criteria: Maximum age** ▾
      - **Maximum eligible age(\*)** : *Not specified*
      - **Unit of age(\*)** : *Not specified*
  - **Eligible gender** : Male, Female
  - **Inclusion criteria** : *Not specified*
  - **Exclusion criteria** : *Not specified*
- **Population of the Project(\*)** ▾
  - **Coverage** : National
  - **Country(ies)** : Germany
  - **Region(s) and/or city(ies)** : *Not specified*

**Groups of diseases or conditions(\*)****Which groups of diseases or conditions were the data collected on?**

× Certain infectious or parasitic diseases (I)

General groups of diseases or conditions on which the data were collected in the Project.

**On which diseases or conditions were the data collected?**

Search ...

Diseases or conditions on which the data were collected in the Project.

**Collects mortality data?**

Search ...

Indication whether mortality data are collected in the Project.

**Administrative information about the Project****Status of the ethics committee approval**

× Request for approval submitted, approval granted

Status of the Project approval from the (leading) ethics committee.

**Overall Project status**

× Completed: Recruitment, data collection, and data quality management completed normally

Overall status of the Project.

**Participants enrolled by invitation?**

Search ...

Specification whether Project participants are selected from a predetermined population.

**Start date**

e.g. January 1, 2015

Start date of data collection for the Project.

**End date**

e.g. January 1, 2015

In case of a Project with patients or other participants, it is the date when the last participant is examined or receives an intervention, or the date when the last participant is followed up.

# Manage Extended Metadata

- List installed EXMD  
[https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types.json](https://ldh.zks.uni-leipzig.de/extended_metadata_types.json)
- Full Read and Write API for EXMD
- Manage EXMD (List / Enable / Disable) [https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types/administer](https://ldh.zks.uni-leipzig.de/extended_metadata_types/administer)

```
ldh.zks.uni-leipzig.de/extended_metadata_types.json
{
  "data": [
    {
      "id": "10",
      "type": "extended_metadata_types",
      "attributes": {
        "title": "MDS 3.0 Interventional Study"
      },
      "links": {
        "self": "/extended_metadata_types/10"
      }
    },
    {
      "id": "11",
      "type": "extended_metadata_types",
      "attributes": {
        "title": "MDS 3.0 Non-Interventional Study"
      },
      "links": {
        "self": "/extended_metadata_types/11"
      }
    }
  ]
}
```

# Manage Extended Metadata

- List installed EXMD  
[https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types.json](https://ldh.zks.uni-leipzig.de/extended_metadata_types.json)
- Full Read and Write API for EXMD
- Manage EXMD (List / Enable / Disable) [https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types/administer](https://ldh.zks.uni-leipzig.de/extended_metadata_types/administer)

```
ldh.zks.uni-leipzig.de/extended_metadata_types.json
{
  "data": [
    {
      "id": "2",
      "type": "projects",
      "attributes": {
        "discussion_links": [],
        "extended_attributes": {
          "extended_metadata_type_id": "108",
          "attribute_map": {
            "Resource_classification_Project": {
              "Resource_classification_type_Project": "Study"
            },
            "Resource_titles_Project": [
              {
                "Resource_titles_text_Project": "A multicentre patients with long-term virologic response",
                "Resource_titles_language_Project": "English"
              }
            ],
            "Resource_acronyms_Project": [
              {
                "Resource_acronyms_text_Project": "STOP-NUC",
                "Resource_acronyms_language_Project": "English"
              }
            ]
          }
        }
      }
    }
  ]
}
```

```
ldh.zks.uni-leipzig.de/projects/2.json
{
  "data": {
    "id": "2",
    "type": "projects",
    "attributes": {
      "discussion_links": [],
      "extended_attributes": {
        "extended_metadata_type_id": "108",
        "attribute_map": {
          "Resource_classification_Project": {
            "Resource_classification_type_Project": "Study"
          },
          "Resource_titles_Project": [
            {
              "Resource_titles_text_Project": "A multicentre patients with long-term virologic response",
              "Resource_titles_language_Project": "English"
            }
          ],
          "Resource_acronyms_Project": [
            {
              "Resource_acronyms_text_Project": "STOP-NUC",
              "Resource_acronyms_language_Project": "English"
            }
          ]
        }
      }
    }
  }
}
```

# Manage Extended Metadata

- List installed EXMD  
[https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types.json](https://ldh.zks.uni-leipzig.de/extended_metadata_types.json)
- Full Read and Write API for EXMD
- Manage EXMD (List / Enable / Disable) [https://ldh.zks.uni-leipzig.de/extended\\_metadata\\_types/administer](https://ldh.zks.uni-leipzig.de/extended_metadata_types/administer)

The image shows a sequence of browser screenshots. The top screenshot displays the JSON response for `ldh.zks.uni-leipzig.de/extended_metadata_types.json`, which is an array of objects under a `"data"` key. The middle screenshot shows `ldh.zks.uni-leipzig.de/projects/2.json`. The bottom screenshot shows the administrator interface at `ldh.zks.uni-leipzig.de/extended_metadata_types/administer`. This interface includes a search bar, a user profile for Dr. Frank Meineke, and a table of metadata types with columns for Internal Id, Title, Supported Type, Number of times used, and Status. Each row has an 'Enable' or 'Disable' button.

Internal Id	Title	Supported Type	Number of times used	Status	Action
18	MIAPPE metadata v1.1	Assay	0	Disabled	Enable
10	MDS 3.0 Interventional Study	Document	0	Disabled	Enable
11	MDS 3.0 Non-Interventional Study	Document	0	Disabled	Enable
14	MDS 3.0 Interventional Study	Investigation	0	Disabled	Enable
15	MDS 3.0 Non-Interventional Study	Investigation	0	Disabled	Enable
16	MIAPPE metadata v1.1	Investigation	0	Disabled	Enable
78	Nfdi4Health MDS 3.2	Investigation	0	Disabled	Enable
168	Nfdi4Health MDS 3.3	Investigation	0	Enabled	Disable
108	Nfdi4Health MDS 3.3	Project	2	Enabled	Disable

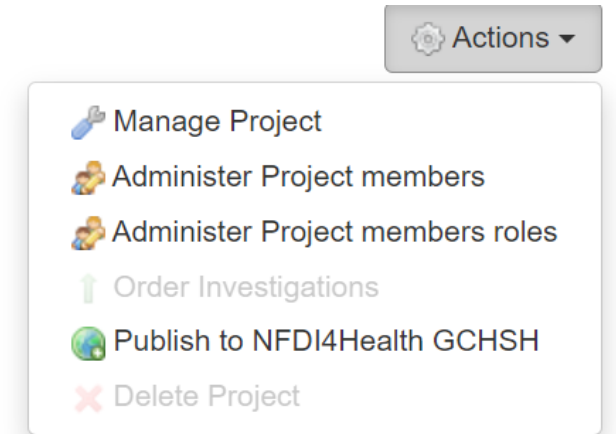
# Further adaptations

Some of our DHO customers (from clinical trials) found ISA not suitable – they preferred the a more simple resource hierarchy based on Programme/Project level.

We strengthened then Project level

- implemented Extended Metadata Attachment Project resource
- implemented using Tags for Project Level

We added the NFDI4Health proprietary feature to publish MDS to the German Health Search Hub.



We like to make Seek terminology to be more relevant to designated communities, e.g. Customize / “Localize” for different Data Holding Organisation, e.g. for Clinical Trial Centers

- “Programme” → “Sponsors”
- “Projects” → “Clinical Trials” (if ISA not used)
- “PALs” → PI (Principal Investigators)

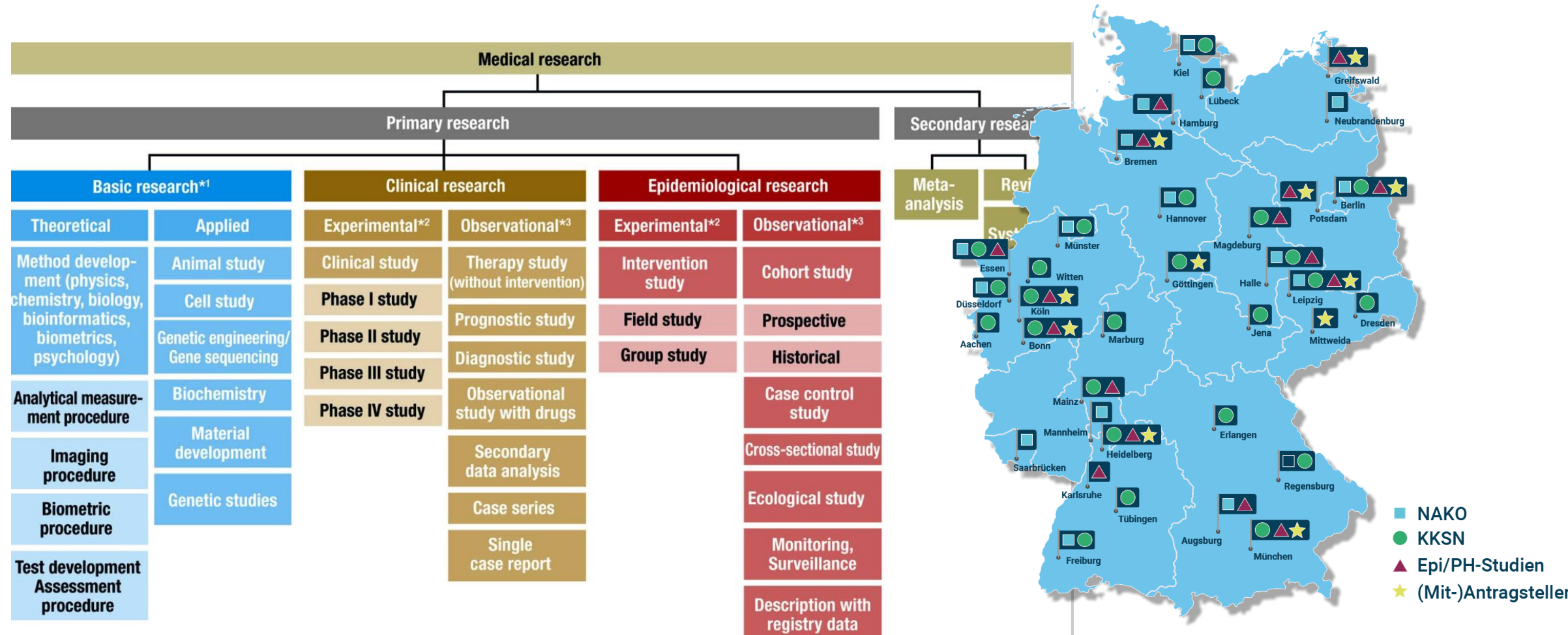
# Outlook / Aim

Make Extended Metadata to be used as simple as other resources

- upload a specification (format to be discussed) or even an UI editor
- attach to resource types via UI
- Full CRUD (Create, read, update and delete) of Extended Metadata filelevel
- versioning with SEEK
- migration support (as far as automatically possible)

→ Make and distribute the FAIRDOM-SEEK Plattform an universal Research Data/Information Management tool for Medical Research

# SEEK is great – we aim to make it usable for Medical Research Applications



Röhrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009 Apr;106(15):262-8.

# Thanks for your Attention!

More Information and Flyer:  
<https://www.nfdi4health.de/service/local-data-hub.html>



## Kontakt:

**Frank Meineke**

[frank.meineke@imise.uni-leipzig.de](mailto:frank.meineke@imise.uni-leipzig.de)

**IMISE,  
University of Leipzig**

**Martin Golebiewski**

[Martin.Golebiewski@h-its.org](mailto:Martin.Golebiewski@h-its.org)

**HITS Heidelberg**

[www.nfdi4health.de](http://www.nfdi4health.de)



UNIVERSITÄT  
LEIPZIG

Heidelberg Institute for  
Theoretical Studies



Gefördert durch

**DFG** Deutsche  
Forschungsgemeinschaft

