



# miracum

Medical Informatics in Research and Care in University Medicine

## Aufbau einer Archivierungsinfrastruktur für Forschungsdaten

**Felix Schönecker**

**07.03.2023**

**MIRACUM-DIFUTURE-Kolloquium**



**hochschule mannheim**





- Die Teilnehmenden kennen:
  - die Bedeutung der Archivierung von Forschungsdaten für die gute wissenschaftliche Praxis
  - die Ziele, das Konzept und die Komponenten der Archivierungsinfrastruktur für die Forschung sowie
- Erläuterungen Berücksichtigung der FAIR-Prinzipien in der Archivierungslösung
- Bedeutung der Metadaten für die Archivierung
- Architektur der Archivierungsinfrastruktur für das MIRACUM-Projekt
- Vorstellung der Komponenten der Archivierungslösung, u. a. FAIRDOM Seek

# Gute wissenschaftliche Praxis



miracum

## Leitlinie 17: Archivierung

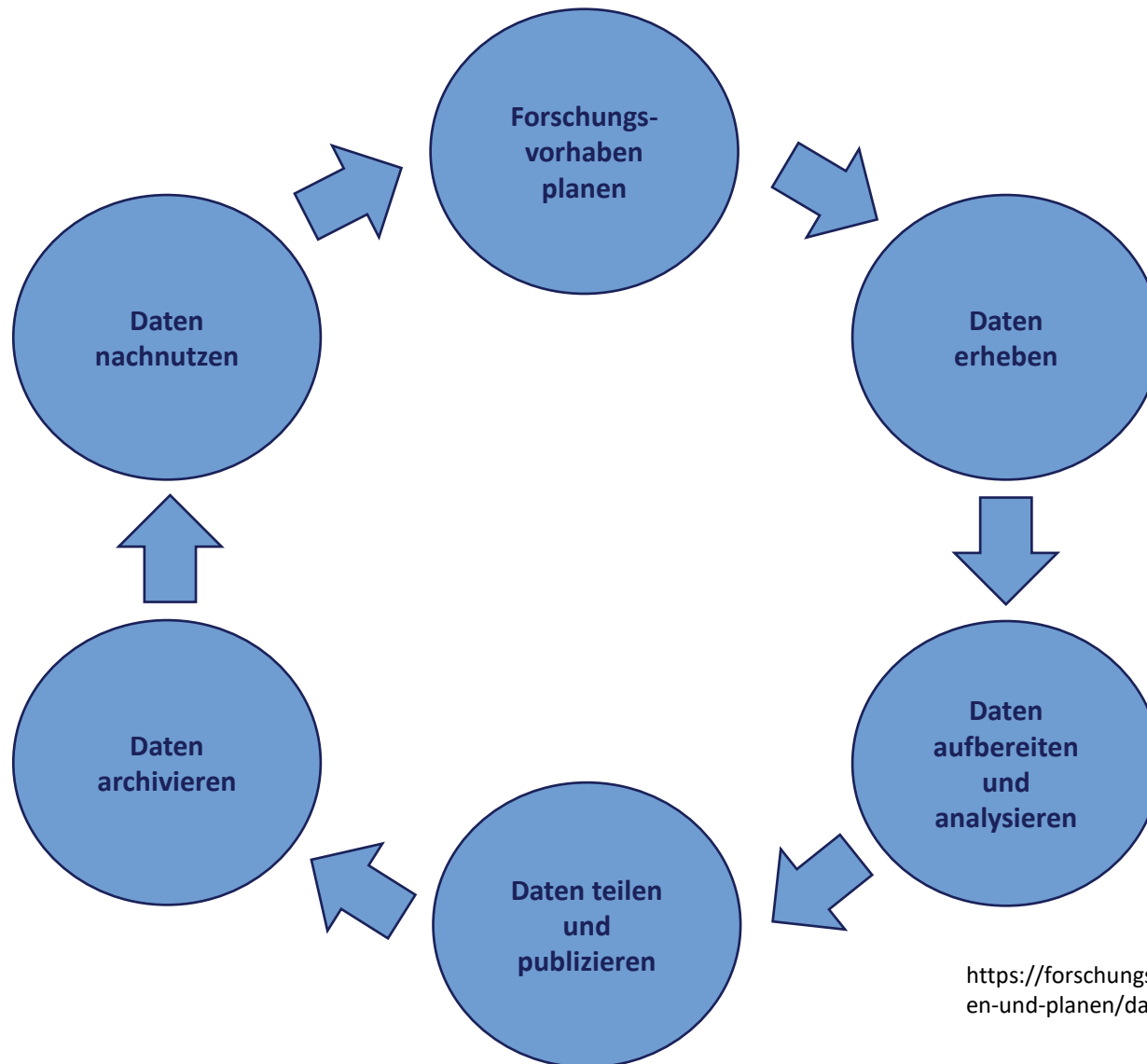
„Wissenschaftlerinnen und Wissenschaftler sichern öffentlich zugänglich gemachte Forschungsdaten beziehungsweise Forschungsergebnisse sowie die ihnen zugrunde liegenden, zentralen Materialien und gegebenenfalls die eingesetzte Forschungssoftware, gemessen an den Standards des betroffenen Fachgebiets, in adäquater Weise und bewahren sie für einen angemessenen Zeitraum auf.“

Quelle: <https://wissenschaftliche-integritaet.de/kodex/archivierung/>

# Datenlebenszyklus



miracum



<https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>

# FAIRe Daten



miracum

**Auffindbar**  
(Findable)



**Zugänglich**  
(Accessible)



**Interoperabel**  
(Interoperable)



**Wiederverwendbar**  
(Reusable)



<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten//>

# Berücksichtigung der FAIR-Prinzipien in der Archivierungslösung (1)



miracum

- Auffindbarkeit

F1. (Meta-) Daten erhalten eine global eindeutige und dauerhafte Kennung.

F2. Daten werden mit umfangreichen Metadaten beschrieben (siehe R1).

F3. Metadaten enthalten eindeutig und explizit die Kennung der von ihnen beschriebenen Daten.

F4. (Meta-) Daten werden in einer durchsuchbaren Ressource registriert oder indiziert.

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

# Berücksichtigung der FAIR-Prinzipien in der Archivierungslösung (2)



miracum

- Zugänglichkeit

A1. (Meta-) Daten können anhand ihrer Kennung unter Verwendung eines standardisierten Kommunikationsprotokolls abgerufen werden.

A1.1 Das Protokoll ist offen, kostenlos und universell implementierbar.

A1.2 Das Protokoll ermöglicht bei Bedarf ein Authentifizierungs- und Autorisierungsverfahren.

A2. Auf Metadaten kann zugegriffen werden, auch wenn die Daten nicht (mehr) verfügbar sind.

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

# Berücksichtigung der FAIR-Prinzipien in der Archivierungslösung (3)



miracum

- Interoperabilität

I1. (Meta-) Daten verwenden eine formale, zugängliche, gemeinsame und allgemein anwendbare Sprache für die Wissensrepräsentation.

I2. (Meta-) Daten verwenden Vokabulare, die den FAIR-Prinzipien folgen.

I3. (Metadaten) enthalten qualifizierte Verweise auf andere (Meta-) Daten.

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>



# Berücksichtigung der FAIR-Prinzipien in der Archivierungslösung (4)



miraculum

- Wiederverwendbarkeit

R1. (Meta-) Daten werden mit einer Vielzahl genauer und relevanter Attribute ausführlich beschrieben.

R1.1. (Meta-) Daten werden mit einer eindeutigen und zugänglichen Datennutzungslizenz veröffentlicht.

R1.2. (Meta-) Daten sind mit detaillierten Informationen über die Entstehung versehen.

R1.3. (Meta-) Daten entsprechen domänenrelevanten Community-Standards.

<https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten/>

# Bedeutung der Metadaten für die Archivierung



miracum

## Geschlecht:

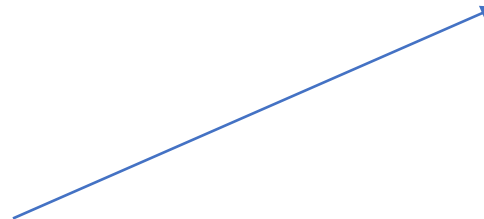
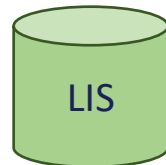
- Männlich
- Weiblich
- Unbekannt

## Sex:

- M
- F

## S:

- 1
- 2



## Spalte 1

2

M

1

Weiblich

2

F

F

# Bedeutung der Metadaten für die Archivierung



miracum

- Standardisierte Metadatenmodelle, -elemente und -kataloge
  - fördern den Austausch von Daten und Informationen zwischen Projektpartnern und IT-Systemen
  - sichern gemeinsame Interpretation der Daten
  - verbessern die Aussagekraft (und Interpretierbarkeit)
- Auffindbarkeit ermöglichen (Schlagworte, DOI)

# Metadatenatz Datenanfrage



miracum

## Request Metadaten

<b>Request Title</b>	Title of the Data Request / Project
<b>Request-ID</b>	Unique Request ID (Based on Jira-Epic ID)
<b>Data Requester</b>	Name of the Requester contact data of the requester (E-mail, Phone)
<b>Affiliation</b>	Affiliation of the Data Requester (Clinic/Department)
<b>Request Date</b>	Date of the sent request
<b>Status</b>	one of the status icons above
<b>Completion Date</b>	Date of the request completion

# Zielsetzung der Archivierungsinfrastruktur



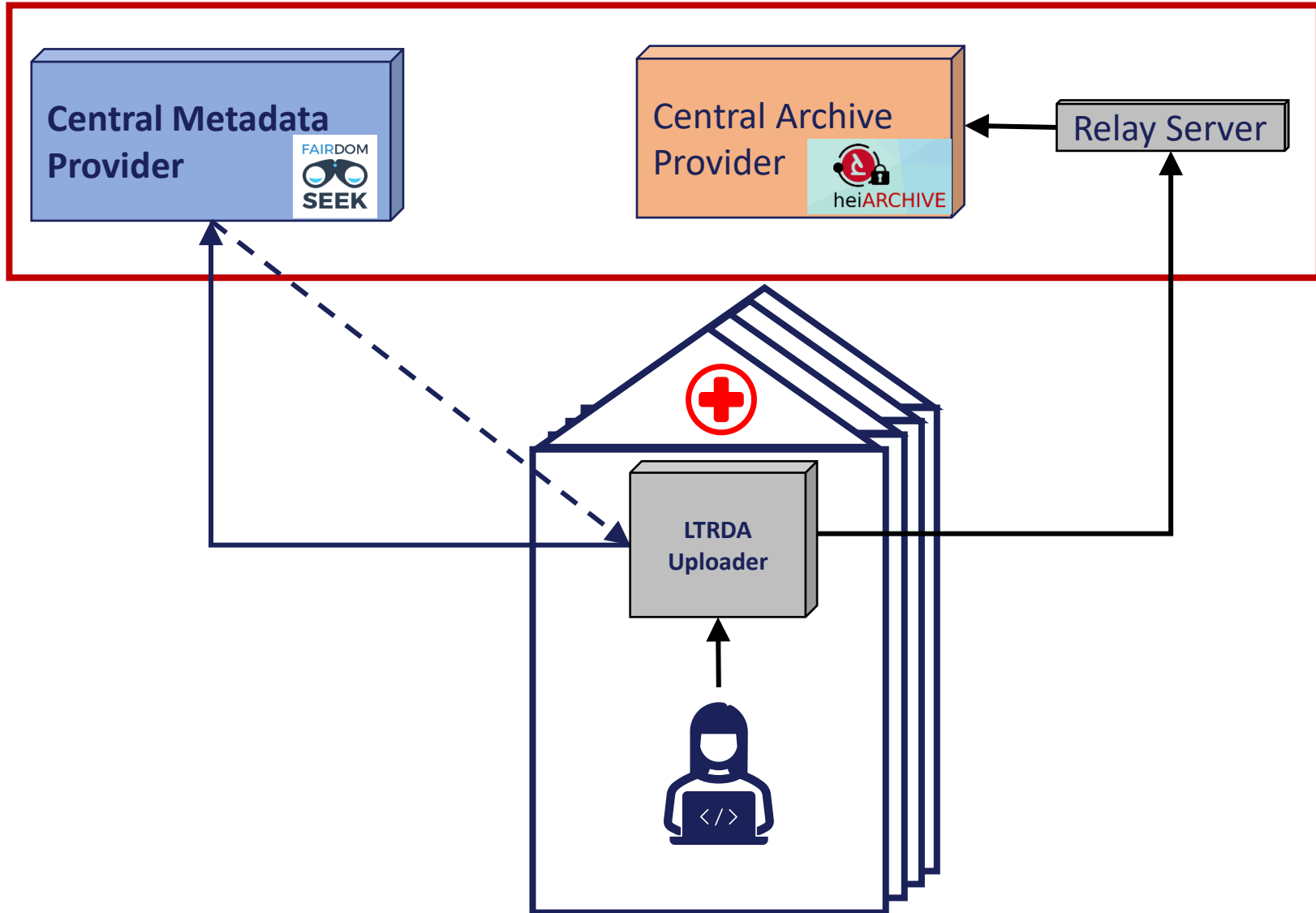
miracum

- Archivierung standortübergreifender Studien und Forschungsergebnisse
- Nachvollziehbarkeit der Ergebnisse anhand der archivierten Daten erhöhen
- Nachnutzung ermöglichen
- Zugriffsmöglichkeiten verwalten

# Architektur der Archivlösung



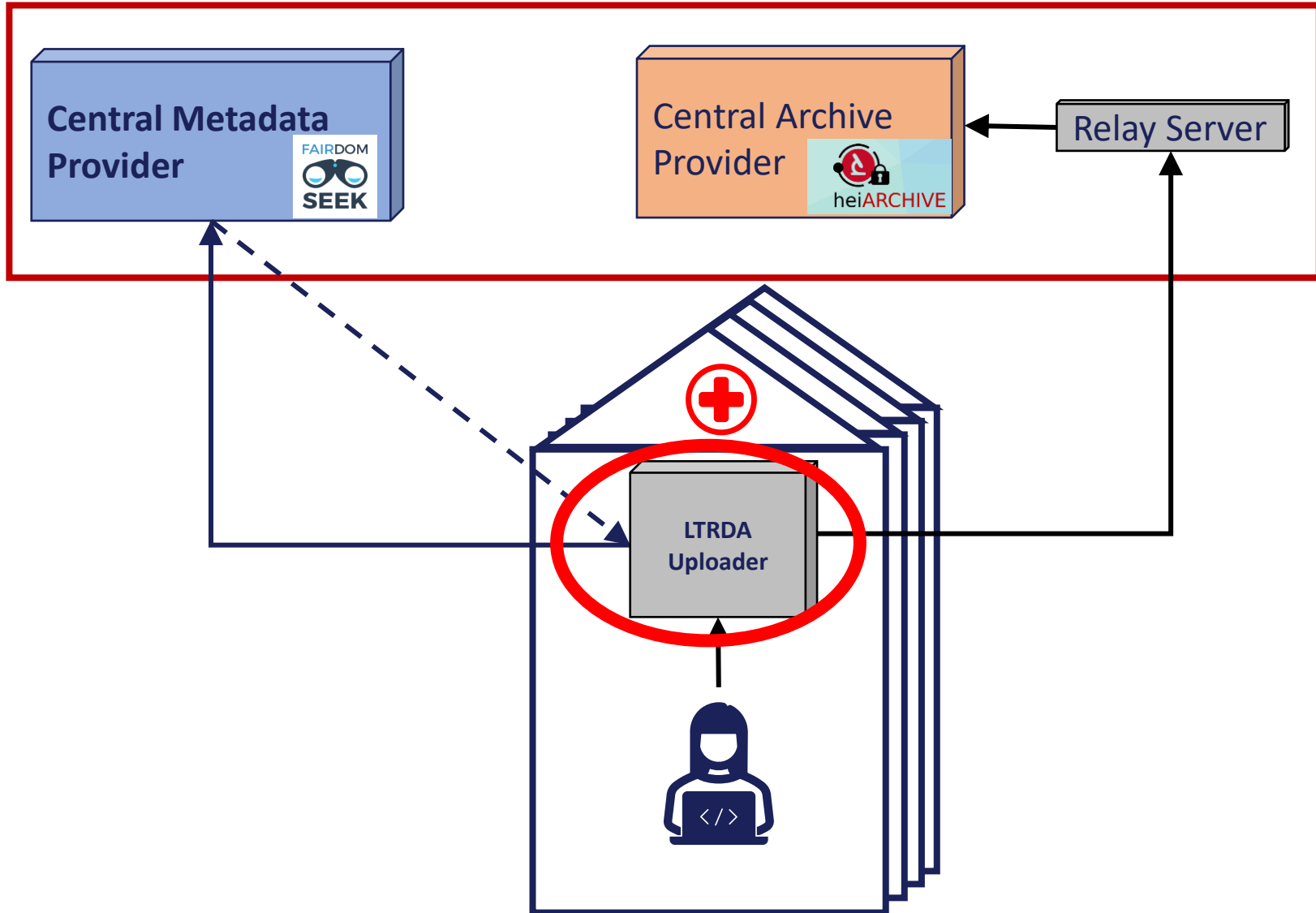
miracum



# LTRDA Uploader



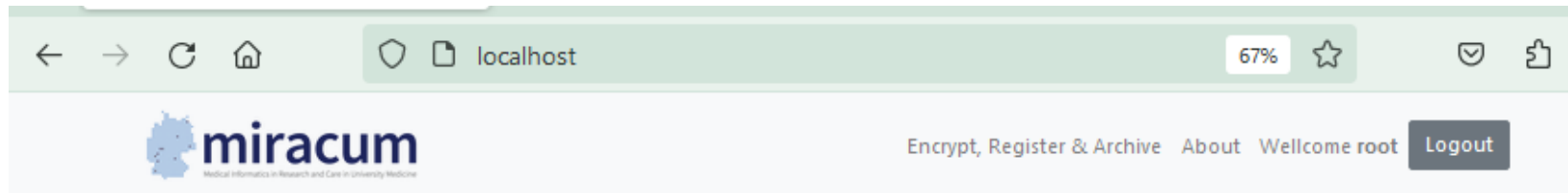
miracum



# LTRDA Uploader (1)



miracum



- Django Webanwendung
- Docker Verteilung
- Übermittlung der Eingaben zum Metadaten Repository





# LTRDA Uploader (2)

Dataset Description

[Files](#)

Title

Title of the dataset

Seek4Science Project

Select a Project



Investigation

Select an Investigation



Study

Select a Study



Assay

Select an Assay



Keywords

Keywords

Insert keywords separated by ";"

Description

Description



Project: responsible person

pro\_responsible



Project: start date

start\_date



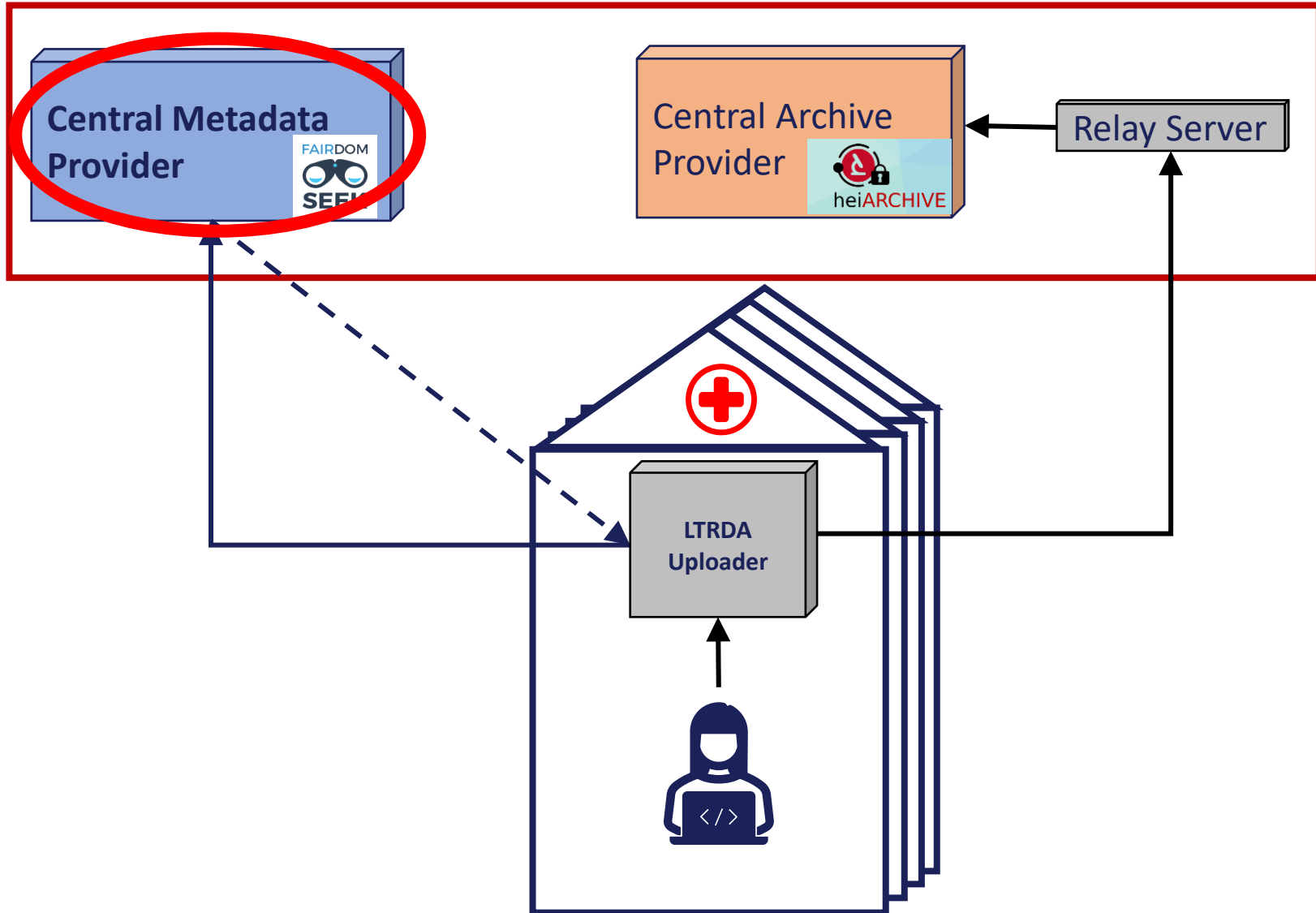
Request: contact requester

contact



Request: responsible Person

# Metadata Provider



miracum



# Was ist FAIRDOM SEEK?

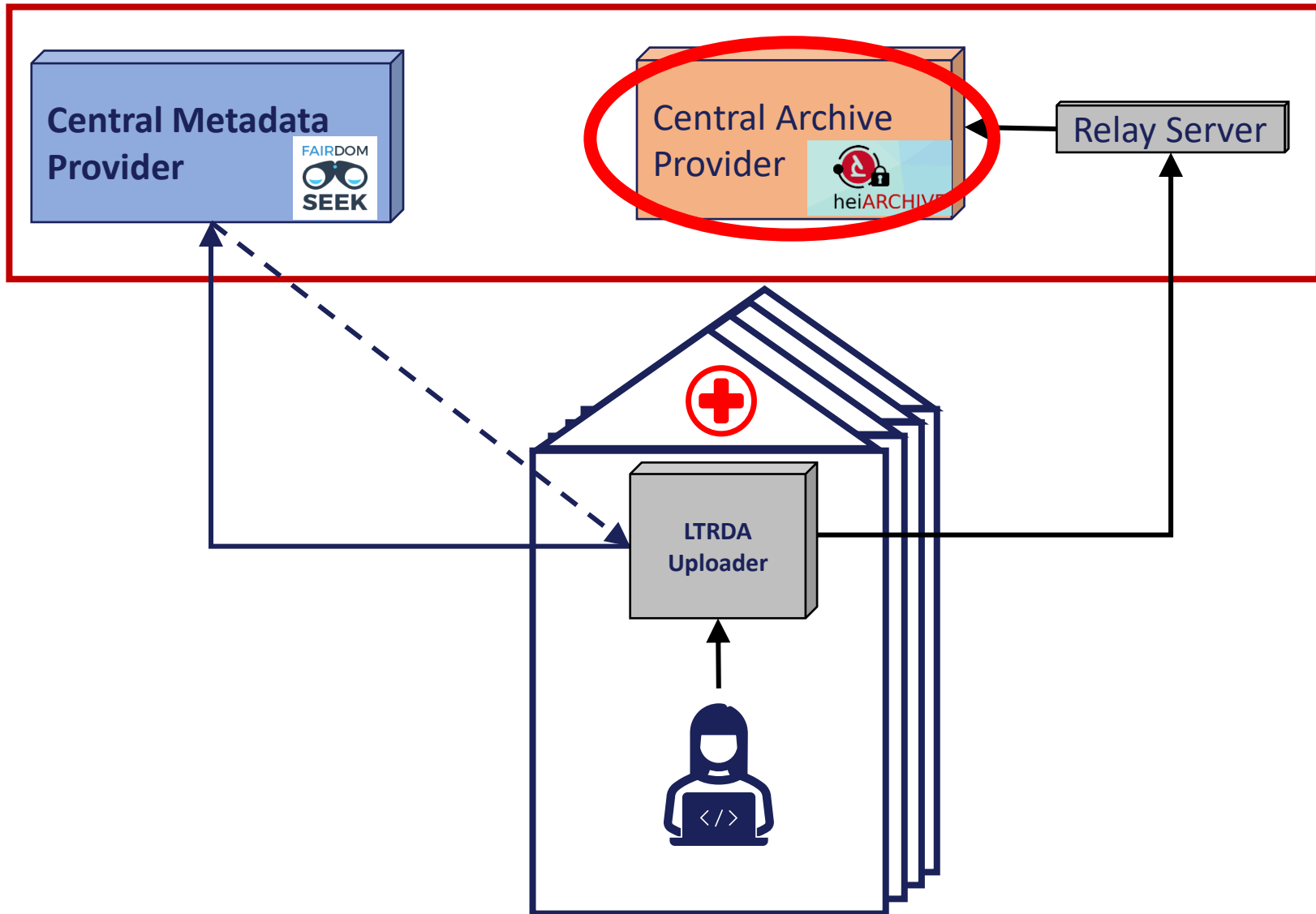
- Plattform für Forschungsdatenmanagement und Archivierung
  - webbasiert
  - Open Source
  - unterstützt Integration
- Archivmaterial: Forschungsdaten, Modelle, Protokolle, Abläufe, Publikationen, Simulationen
- Organisieren, Teilen, Veröffentlichen, Vernetzen
- FAIRDOM Hub, Projekt Hub, Lokale Hub



# Archive Provider



miracum



- zentraler Datenarchivierungsdienst heiARCHIVE der Universität Heidelberg
- Betrieb durch Kompetenzzentrum Forschungsdaten
- Archivvarianten werden angeboten:
  - Vergabe der Identifier
  - Anlage von Kopien
  - Georeplikation
- Relay Server für Verbindung zwischen LTRDA Server und heiArchive
- derzeit im Pilotbetrieb



miracum



# Vielen Dank für Ihre Aufmerksamkeit



@miracum\_de



hochschule mannheim



GEFÖRDERT VOM

Bundesministerium  
für Bildung  
und Forschung