



# Qualitätsmanagement und Aktualität von Daten

Berlin, 17. Januar 2024  
Stiftung Deutsche Kinemathek



---

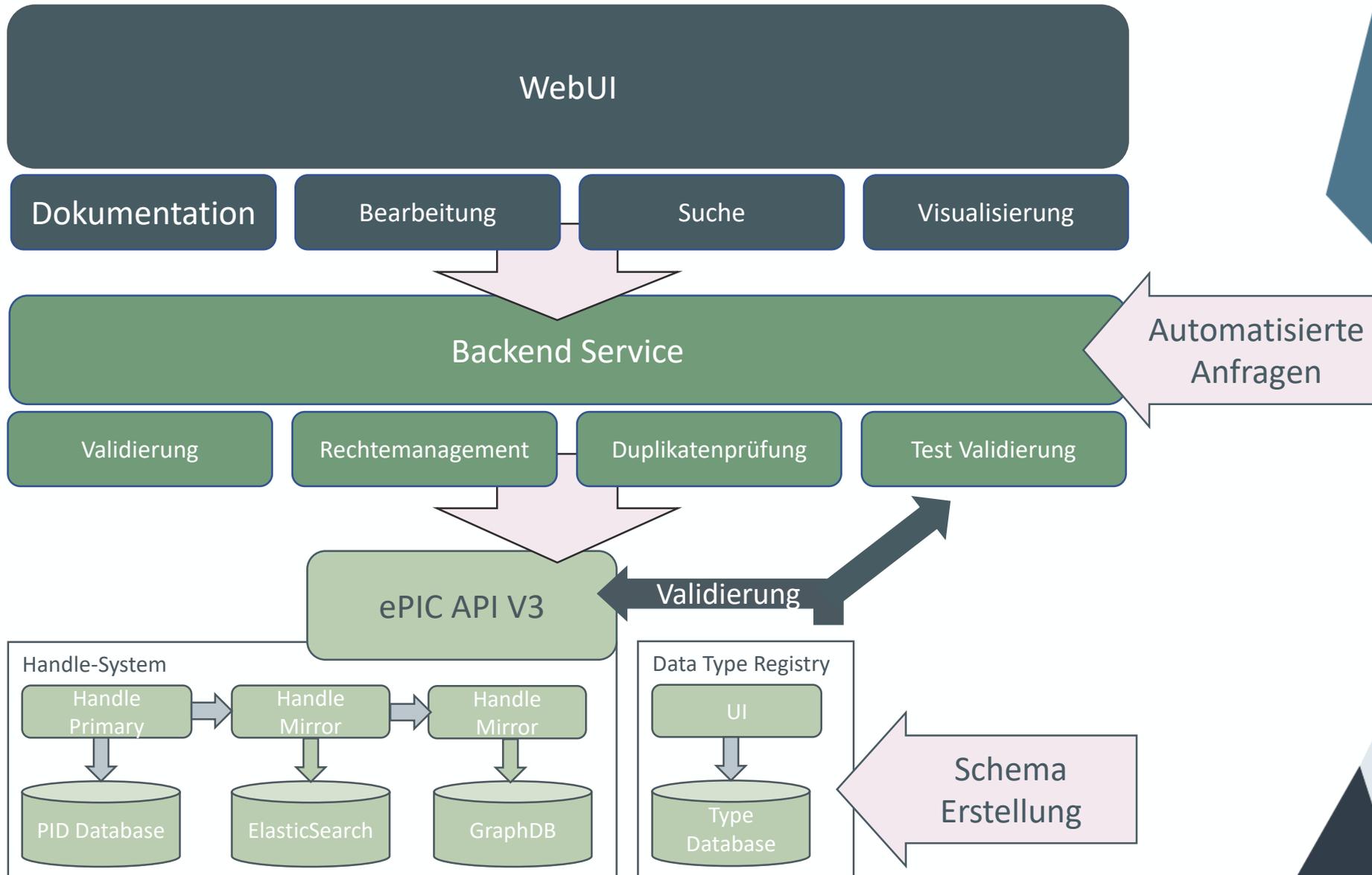
# Qualitätsmanagement

Definition Qualität nach DIN EN ISO 9000:2015-11 als

„Grad, in dem ein Satz inhärenter Merkmale eines Objekts  
Anforderungen erfüllt“



# Architektur

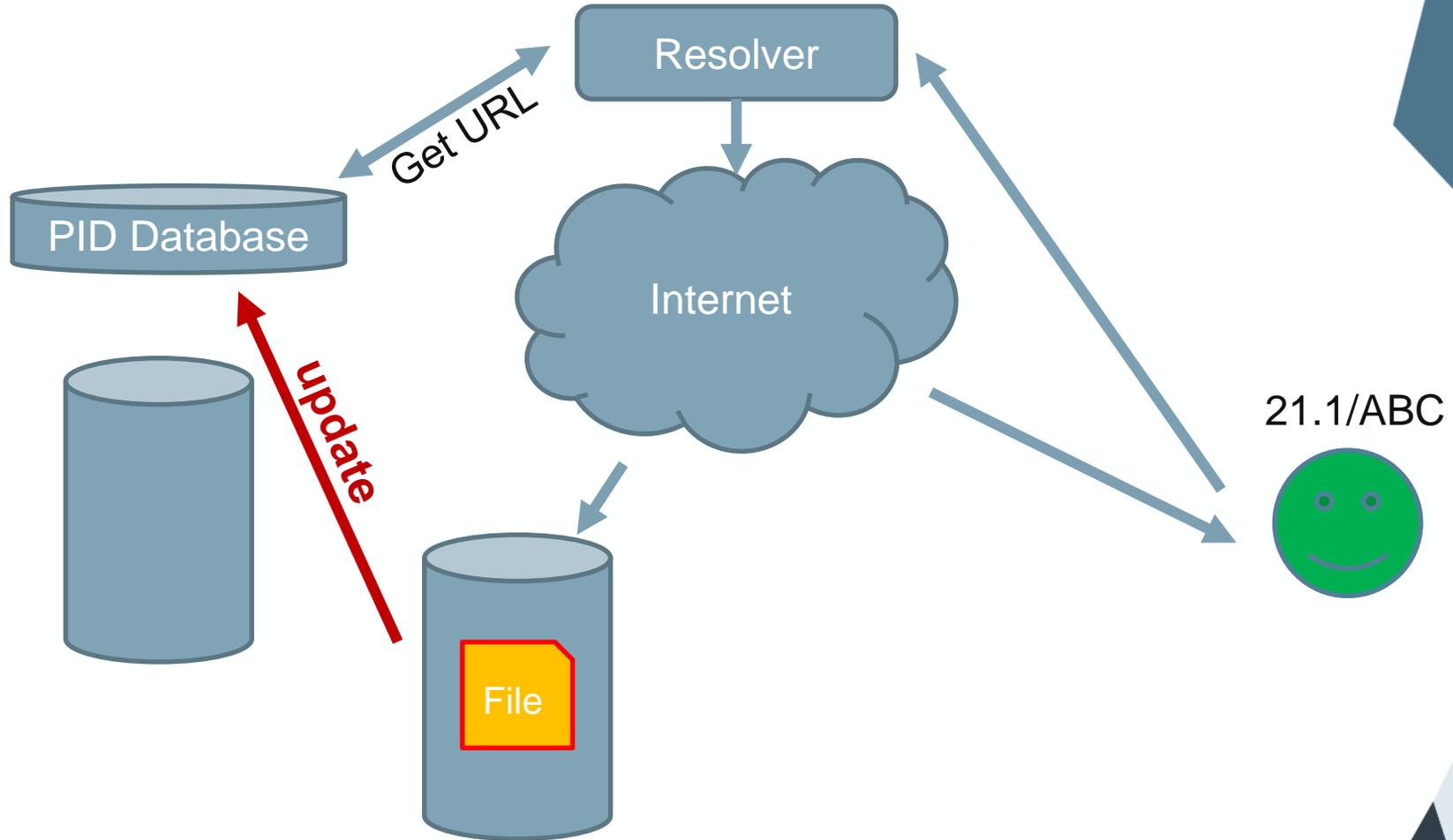


## Welches PID System?

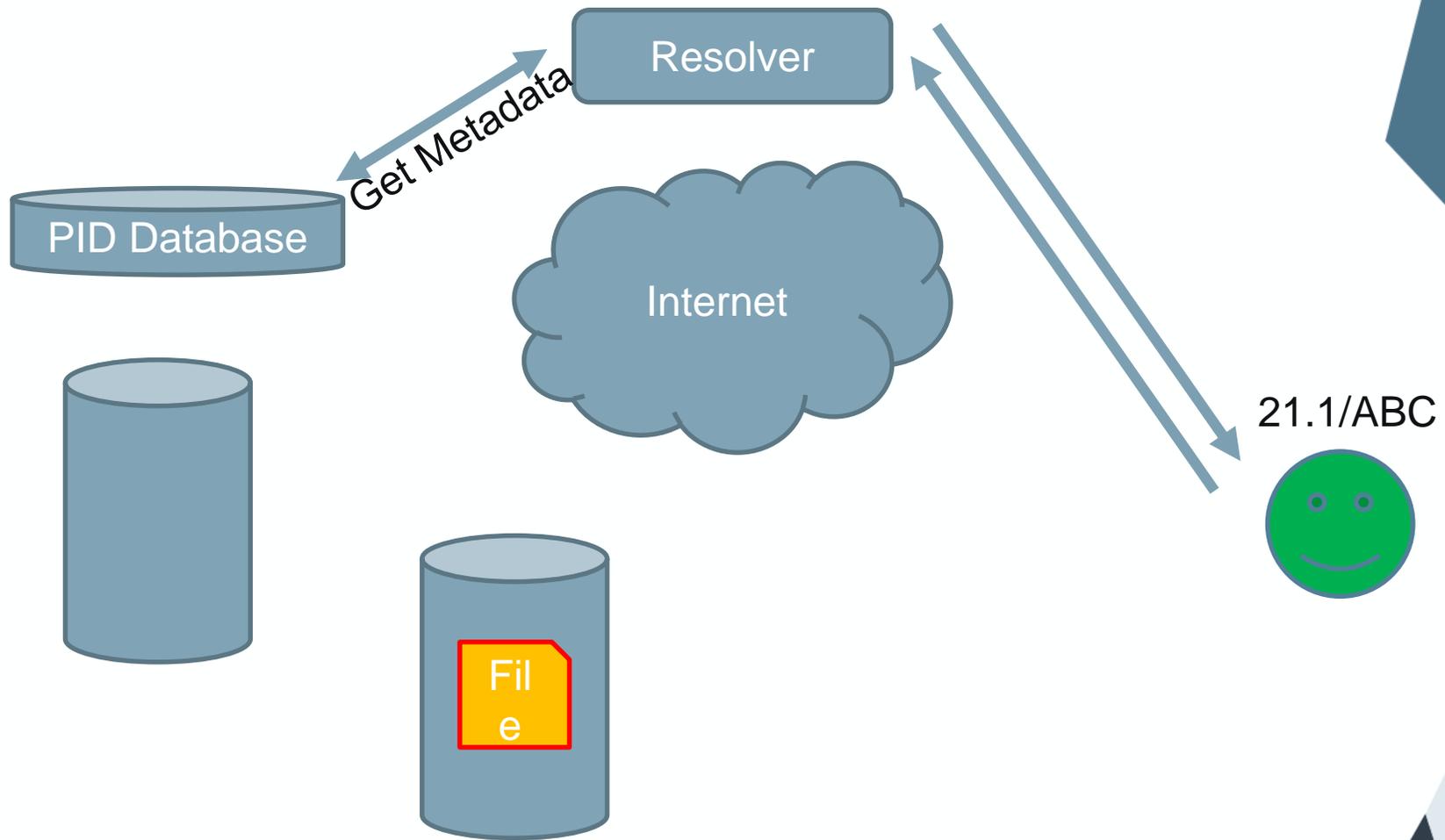
- Viele unterschiedliche Möglichkeiten ARK, URN, Handle-basiert
- Handles (DOI, ePIC, RAIDs) sind global auflösbar
- Handles enthalten Metadaten
  - DOI : URL
  - ePIC: offen

ePICs als Grundbaustein für das Verbundsystem!

# Auflösung zur Ressource



# Auflösung zu den Metadaten



# Metadatenschemata

# ROR

# Data Cite

```
<xs:schema targetNamespace="http://datacite.org/schem
<xs:import namespace="http://www.w3.org/XML/199
<xs:include schemaLocation="include/datacite-titleTy
<xs:include schemaLocation="include/datacite-contrib
<xs:include schemaLocation="include/datacite-dateTy
<xs:include schemaLocation="include/datacite-resourc
<xs:include schemaLocation="include/datacite-relatio
<xs:include schemaLocation="include/datacite-related
<xs:include schemaLocation="include/datacite-funderl
<xs:include schemaLocation="include/datacite-descrip
<xs:include schemaLocation="include/datacite-nameT
<xs:include schemaLocation="include/datacite-numbe
+<xs:element name="resource"></xs:element>
<!-- TYPE DECLARATIONS -->
<!-- defines value for mandatory fields -->
+<xs:simpleType name="nonemptycontentStringType">
<!-- definition for nameIdentifier -->
+<xs:complexType name="nameIdentifier"></xs:compl
+<!-->
+<xs:simpleType name="edtf"></xs:simpleType>
<!-- definition for affiliation -->
+<xs:complexType name="affiliation"></xs:complexTy
<!-- defines the value for a year -->
+<xs:simpleType name="yearType"></xs:simpleType>
<!-- definitions for geoLocation -->
+<xs:complexType name="point"></xs:complexType>
+<xs:complexType name="box"></xs:complexType>
+<xs:simpleType name="longitudeType"></xs:simpleT
+<xs:simpleType name="latitudeType"></xs:simpleTyp
</xs:schema>
```

JSON

- \$schema : "http://json-schema.org/schema#"
- type : "object"
- additionalProperties : false

properties

- ip\_addresses
- aliases
- acronyms
- links
- country
- name
- wikipedia\_url
- addresses
- types
- external\_ids
- established
- relationships
- email\_address
- id
- labels
- status

definitions

- required

# PIDINST

ID	Property	Obligation	Occ	Definition
1	Identifier	M	1	Unique string that identifies the instrument instance
1.1	identifierType	M	1	Type of the identifier
2	SchemaVersion	M	1	Version number of the PIDINST schema used in this record
3	LandingPage	M	1	A landing page that the identifier resolves to
4	Name	M	1	Name by which the instrument instance is known
5	Owner	M	1-n	Institution(s) responsible for the management of the instrument. This may include the legal owner, the operator, or an institute providing access to the instrument.
5.1	ownerName	M	1	Full name of the owner
5.2	ownerContact	O	0-1	Contact address of the owner
5.3	ownerIdentifier	O	0-1	Identifier used to identify the owner
5.3.1	ownerIdentifierType	O	1	Type of the identifier
6	Manufacturer	M	1-n	The instrument's manufacturer(s) or developer. This may also be the owner for custom build instruments

NEU: <https://github.com/AV-EFI/av-efi-schema>

## Attribute/Typen

**Schema X.X**

Publikationsjahr

z.B. 2022 -> Integer 4  
Digits

Farbe

z.B. Blau -> String oder  
ENUM Liste?

Datum

z.B. Sept. 14th 2022 ->  
Konflikt mit ISO 8601

Institut

## Attribute/Typen

**Schema X.X**

Publikationsjahr

Farbe

Datum

Institut

---

Komplexe Struktur mit

- Titel
- Adresse
- Leitung
- ...

# Data Type Registry (DTR)



Data Type Registry

AVefi

## Registrierung eines Attributes/Typs

1. Name des Typs
2. Beschreibung
3. Provenienz
4. Versionierung
5. Validierungsschema zur Überprüfung der Werte
6. Persistent Identifier zum Referenzieren des registrierten Objekts
7. Häufigkeiten, Abhängigkeiten und Verpflichtungen

## Vereinfachtes Beispiel

Schem a X.X	Wert
21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWDG, Burckhar dtweg 4}

- Beispiel Datensatz
- Ablageort beliebig

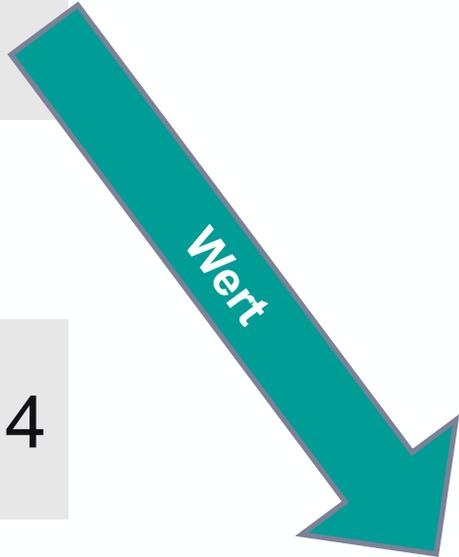
PID auf die Datentyp-Definition

### Vereinfachtes Beispiel

Schem a X.X	Wert
21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWDG, Burckhar dtweg 4}



Data Type  
Registry



Validierung

## PID oder Datenbank PID auf Handle-Basis

URL	gwdg.de
21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWDG, Burckhardt weg 4}

## PID

URL	gwdg.de
-----	---------



21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWDG, Burckhardt weg 4}



## Kernel Information Profiles (KIP)

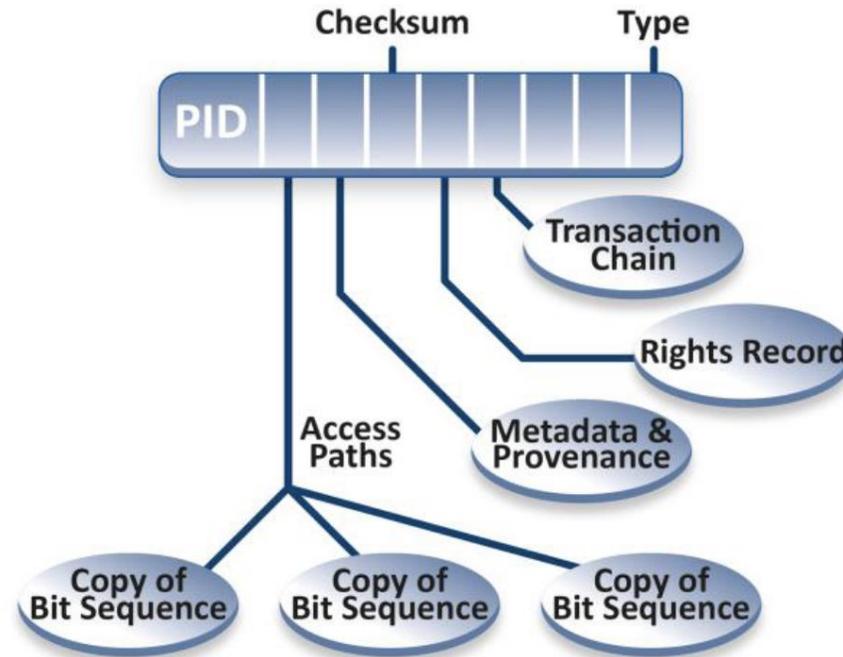
- bestehen aus einer Liste aus Datentypen
- definieren das „Schema“ einer PID
- beinhalten Kernelement zum Verständnis des referenzierten Objekts (z.B. MIME Type)

Anwendungsfälle wie z.B. PIDInst WG : PIDs für Instrumente

## Fair Digital Object Forum

<https://fairdo.org>

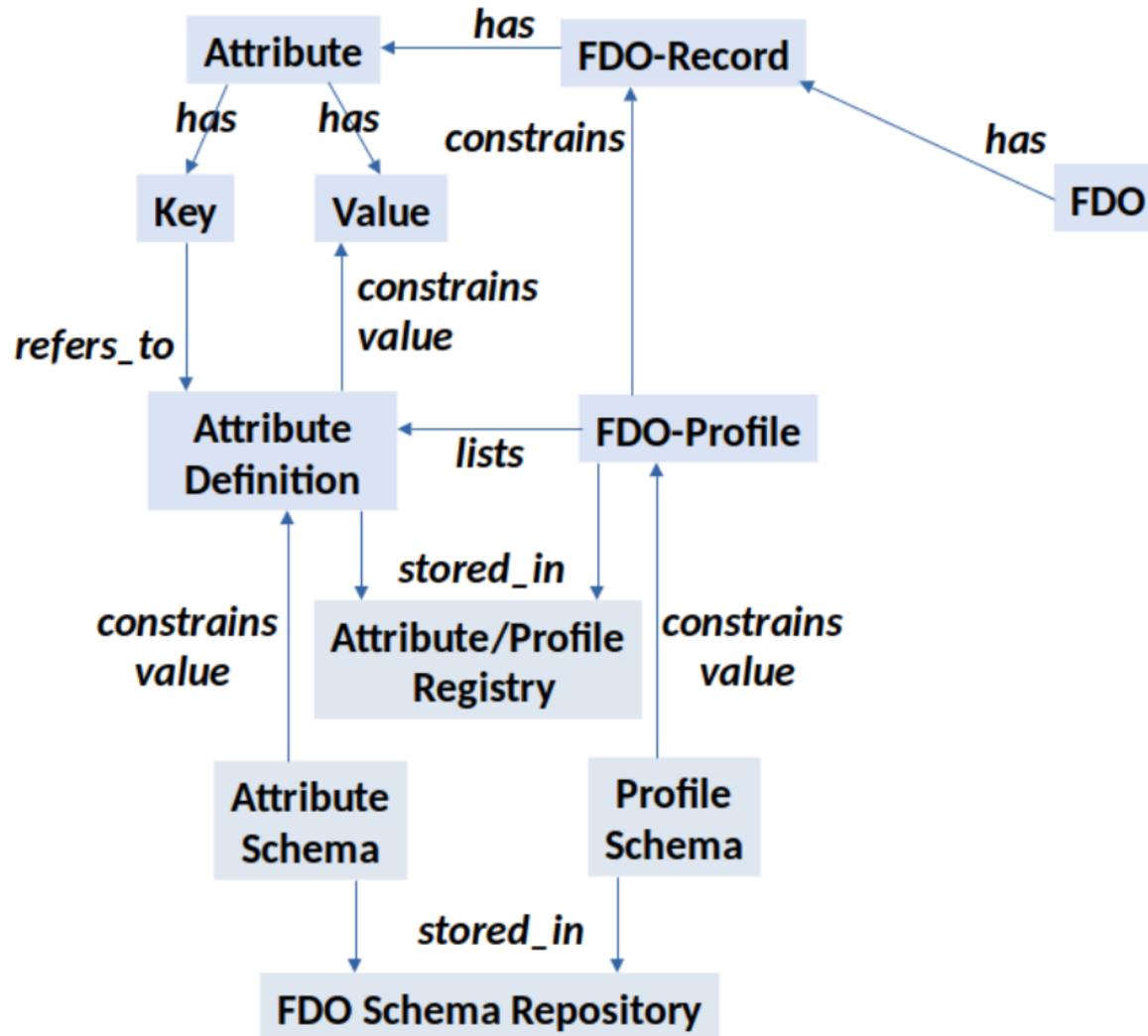
- Creating Fair Digital Objects (FDO)
- Dedicated Digital Object Interface Protocol (DOIP)
- Machine actionable identifiers including KIPs registered in DTR



*Figure 4 indicates the use of the PID record to store immediately relevant metadata and to do the binding of other entities to make DOs FAIR.*

Wittenburg, P.; Strawn, G.; Mons, B.; Bonino, L.; Schultes, E. Digital Objects as Drivers towards Convergence in Data Infrastructures; EUDAT: Helsinki, Finland, 2019.

# FDO (Auszug aus Gesamtkonzept)



# FAIRCORE4EOSC in a nutshell

The project

**Call title:** Deploying EOSC-Core components for FAIR Research and Innovation Action

**Budget:** 10 million EUR

**Duration:** June 2022 – May 2025

**Consortium:** 22 partners, coordinated by CSC – IT Center for Science

**Website:** [faircore4eosc.eu](https://faircore4eosc.eu)

**Key results:** In response to the gaps identified in the SRIA, the project will develop nine new EOSC-Core components aimed to improve the discoverability and interoperability of an increased amount of research outputs.

Presentation 06 | Home Screen



# The 9 FAIRCORE4EOSC components



**EOSC Research Discovery Graph (RDGraph)** to deliver advanced discovery tools across EOSC resources and communities.



**EOSC PID Graph (PIDGraph)** to improve the way of interlinking research entities across domains and data sources on the basis of PIDs.



**EOSC Metadata Schema and Crosswalk Registry (MSCR)** to support publishing, discovery and access of metadata schemas and provide functions to operationalise metadata conversions by combining crosswalks.



**EOSC Data Type Registry (DTR)** to provide user friendly APIs for metadata imports and access to different data types and metadata mappings.



**EOSC PID Meta Resolver (PIDMR)** to offer users a single PID resolving API in which any kind of PID can be resolved through a single, scalable PID resolving infrastructure.



**EOSC Compliance Assessment Toolkit (CAT)** to support the EOSC PID policy compliance and implementation.



**EOSC Research Activity Identifier Service (RAiD)** to mint PIDs for research projects, allowing to manage and track project related activities.

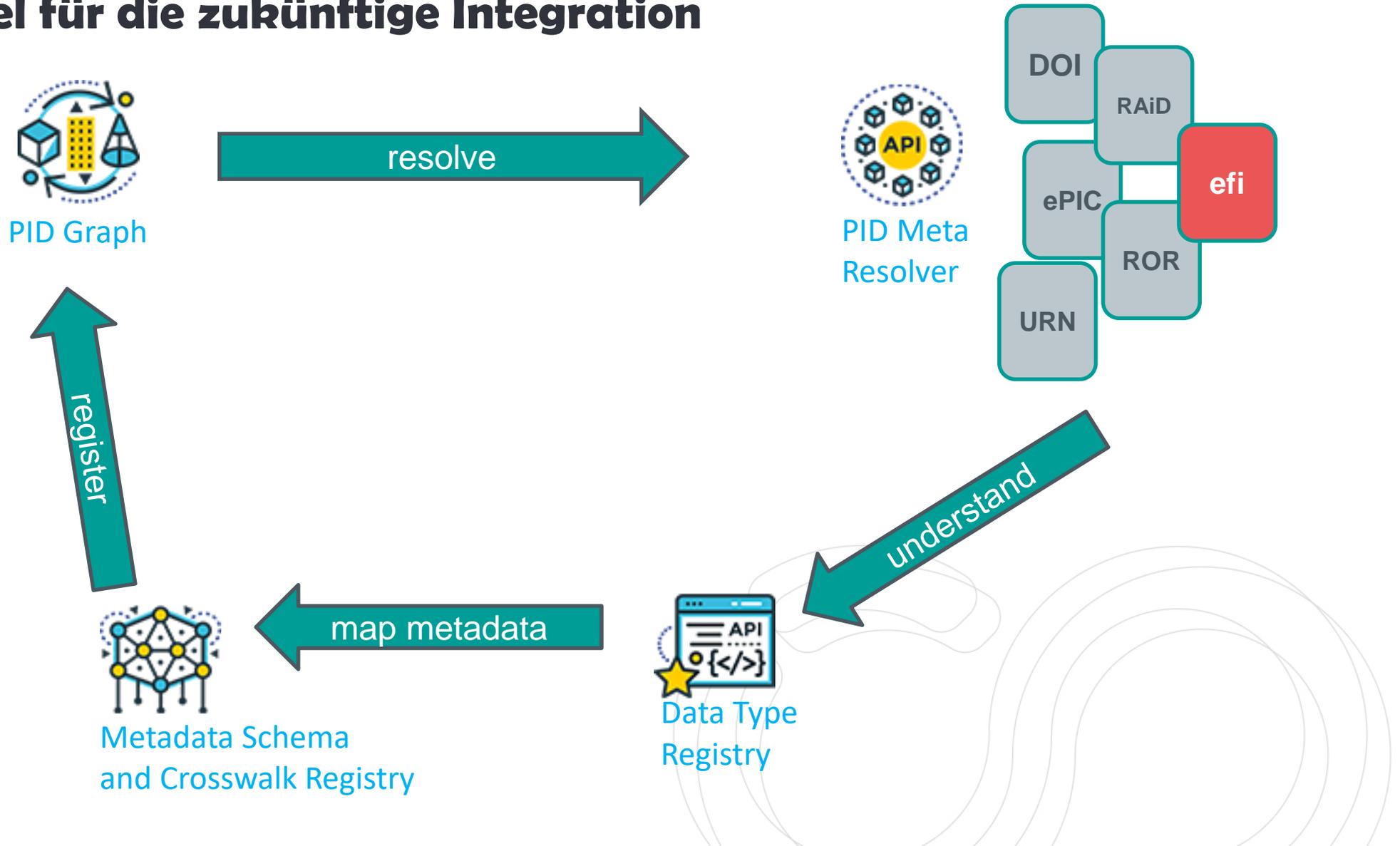


**EOSC Research Software APIs and Connectors (RSAC)** to ensure the long-term preservation of research software in different disciplines.

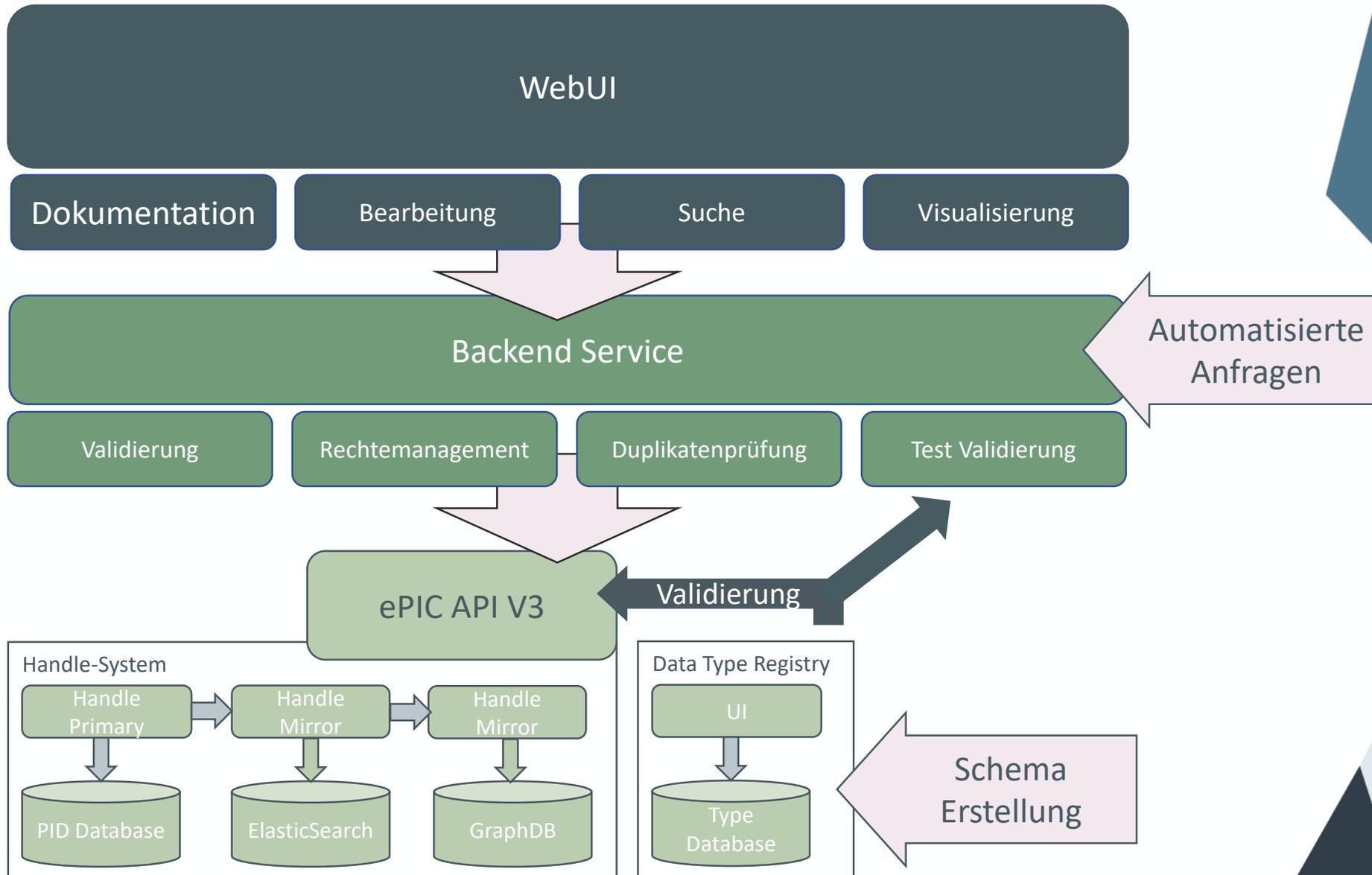


**EOSC Software Heritage Mirror (SWHM)** to equip EOSC with a mirror of the Software Heritage universal source code archive.

# Ein Beispiel für die zukünftige Integration



# Architektur

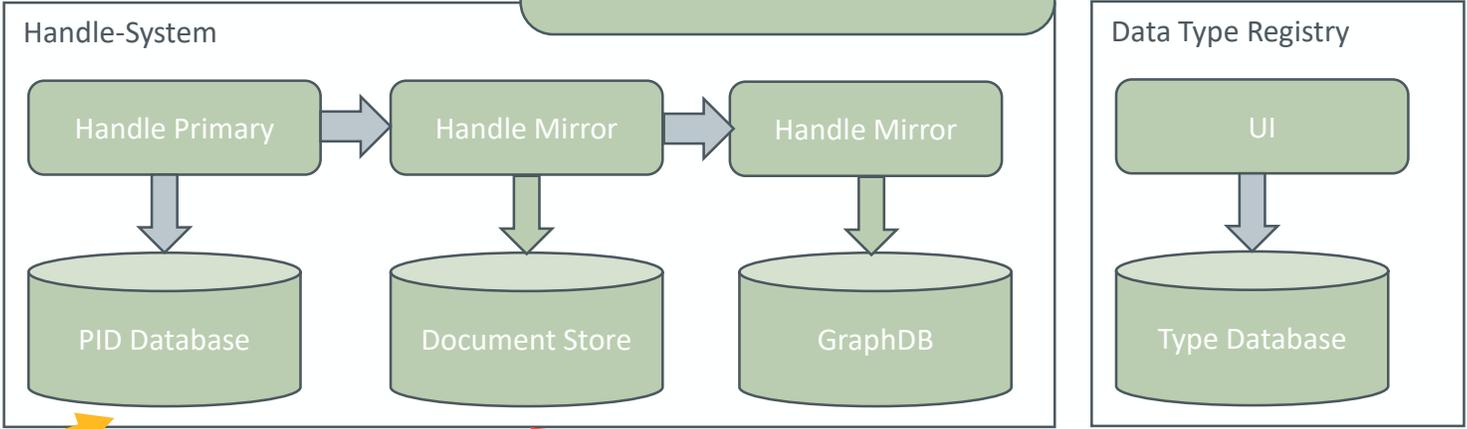
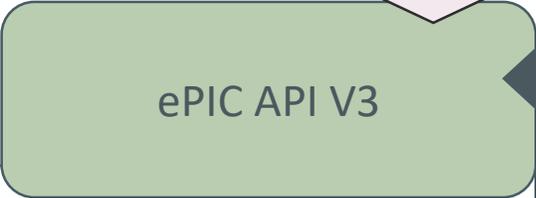
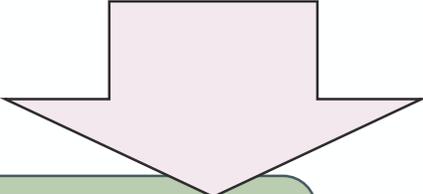


## Architektur Anforderung

- Langfristige Nachhaltigkeit
- Einfache Wartbarkeit
- Geringe Kosten
- Microservice Architektur
- Notwendige Schnittstellen für alle Interessierten

# Datenhaltung

Annahme: Grunddaten liegen alle im PID System  
 • inklusive der Versionen



Grunddaten  
 Persistent in  
 ePIC  
 gesichert



austauschbar



## Diskussionspunkte

- Alles Einträge werden direkt im PID System gesichert
- Korrekturen werden im PID System vorgenommen
- Recherche Oberfläche kann mittels UI unterstützen
- Rechte Management im PID System
  - Werk sind offen: alle können schreiben
  - Versionen und Items „gehören“ Instituten (Schreibrechte)
  - Erfassung der Szenarien und Anforderungen für das Rechte Management

## MEHR INFORMATIONEN

<https://wiki.tib.eu/confluence/pages/viewpage.action?pageId=257984794>

### Kontaktdaten

Sven Bingert: [sven.bingert@gwdg.de](mailto:sven.bingert@gwdg.de)

Elias Oltmanns: [elias.oltmanns@gwdg.de](mailto:elias.oltmanns@gwdg.de)



Creative Commons Namensnennung 3.0 Deutschland  
<http://creativecommons.org/licenses/by/3.0/de>