

# ePIC-Dienste zur Nutzung und Standardisierung von PIDs

Persistent Identifier für FAIRe Forschungsdaten

Sven Bingert (GWDG)  
16. Februar 2023  
RDA Deutschland 2023

# Metadatenschemata



ROR

PIDINST

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-relator
<xs:include schemaLocation="include/datacite-related
<xs:include schemaLocation="include/datacite-funder
<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

ID	Property	Obligation	Occ	Definition
1	Identifier	M	1	Unique string that identifies 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 management of the instrument may include the legal owner, 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/developer. This may also be used for custom build instrument

# Metadatenschemata

- Große Anzahl an Schemata
- Verschiedene Formate
- Oft nicht maschinenlesbar
- Community spezifisch oder generisch
- Fairness

## Benötigt ist

- Mapping zwischen Schemata notwendig
- Validierung von Werten bei gegebenem Schema

# Standards existieren

Wir haben Standards für verschiedene Objekte  
Wie z.B. *Data Cite Schema*, *PID4Inst* oder andere  
heute diskutierte Schemata.

Aber was ist mit der Vielfalt in der Wissenschaft?  
Wie können wir (digitale) Datenobjekte im  
Lebenszyklus mit PIDs begleiten?

Benötigt ist die Möglichkeit *Datentypen neu zu definieren* und zur einer *Beschreibung des referenzierten Objekts* zusammen zu fügen.

# Attribute/Typen



Schema X.X
Publikationsjahr
Farbe
Datum
Institut

z.B. 2022 -> Integer 4 Digits

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

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

# Attribute/Typen

Schema X.X
Publikationsjahr
Farbe
Datum
Institut

Komplexe Struktur mit

- Titel
- Adresse
- Leitung
- ...

# Data Type Registry (DTR)



## ■ Registrierung eines Attributes/Typs

- Name des Typs
- Beschreibung
- Provenienz
- Versionierung
- Validierungsschema zur Überprüfung der Werte
- Persistent Identifier zum Referenzieren des registrierten Objekts
- Häufigkeiten, Abhängigkeiten und Verpflichtungen



Data Type Registry

# Vereinfachtes Beispiel



Schema X.X	Wert
21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWWDG, Burckhardt weg 4}

- Beispiel Datensatz
- Ablageort beliebig



PID auf die Datentyp-Definition



# Vereinfachtes Beispiel

Schema X.X	Wert
21.123	2022
21.124	Blau
21.125	20220914
21.126	{GWWDG, Burckhardt weg 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. MIMEType)
- Anwendungsfälle wie z.B. PIDInst WG : PIDs für Instrumente

# AV-EFI



## Proposal

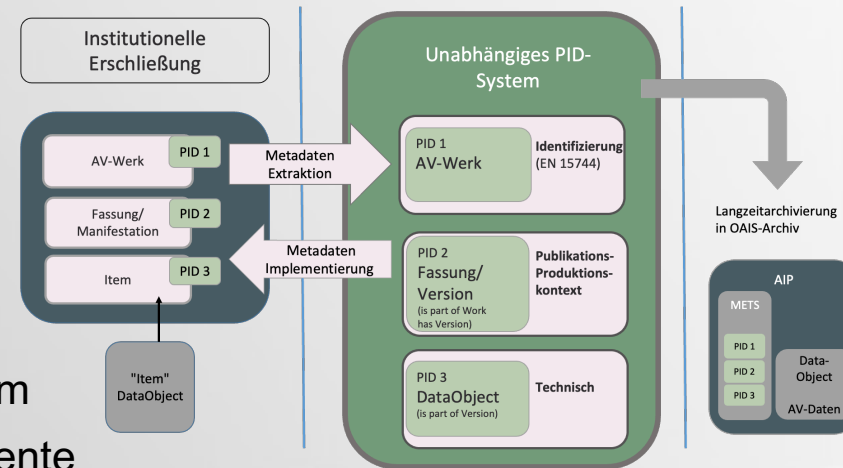
**Automatisiertes Verbundsystem für audiovisuelle Bestände über einheitliche Filmidentifikatoren**

## Ziele

- institutionsübergreifendes, webbasiertes System
- Persistent Identifier (PID) als zentrale Komponente
- Forschungsinfrastruktur zur Identifikation von AV-Werken, den dazugehörigen Fassungen und Datenobjekten
- Strukturierte Erfassung existierender Metadaten

## Partner

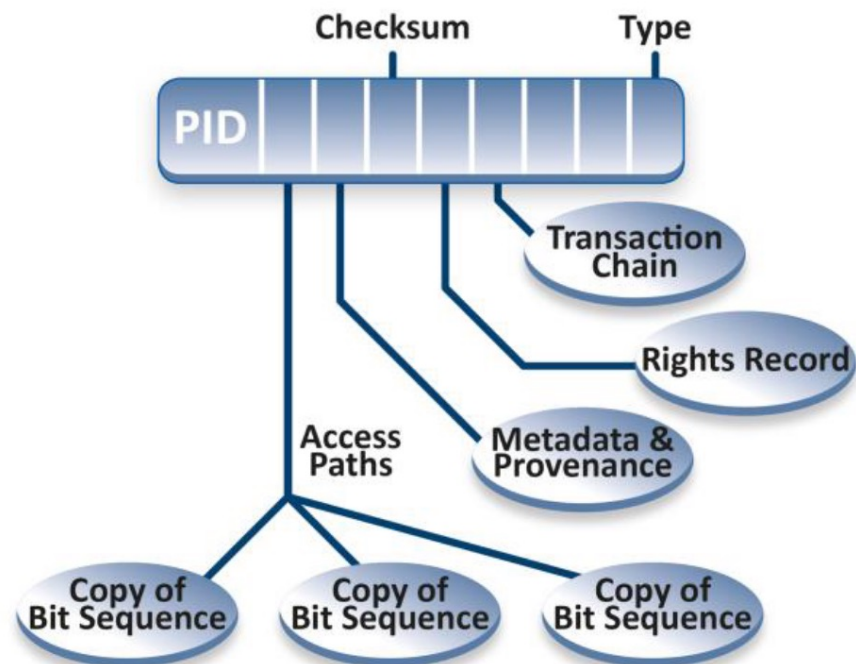
- TIB - Leibniz-Informationszentrum für Technik und Naturwissenschaften
- Stiftung Deutsche Kinemathek – Museum für Film und Fernsehen (SDK)
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDDG)
- Filmmuseum der Landeshauptstadt Düsseldorf (FMDU)



# 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.

# 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.



# 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.

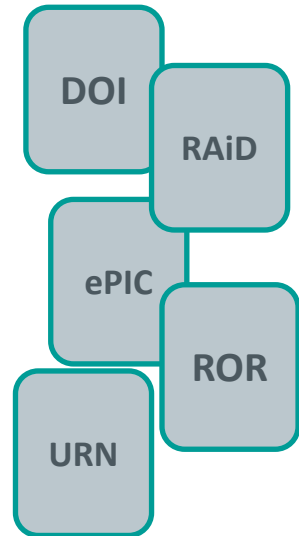


**EOSC Research Software APIs and Connectors (RSAC)** to ensure the long-term preservation of research software in different disciplines.

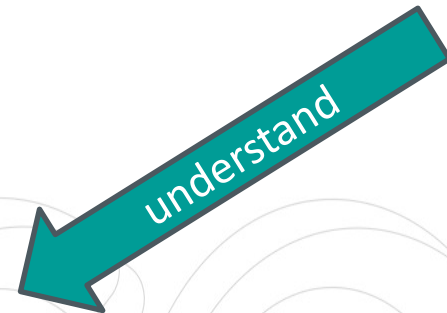
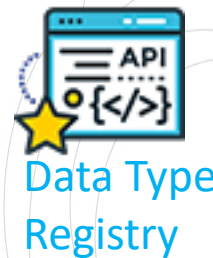


**EOSC Software Heritage Mirror (SWHM)** to equip EOSC with a mirror of the Software Heritage universal source code archive.

# Ein Beispiel für die zukünftige Integration



Metadata Schema  
and Crosswalk Registry





**VIELE DANK FÜR DIE  
AUFMERKSAMKEIT**