A basic Helmholtz Kernel Information Profile for machine-actionable FAIR Digital Objects

Thomas Jejkal, Andreas Pfeil, Jan Schweikert, Anton Pirogov, Pedro Videgain Barranco, Florian Krebs, Christian Koch, Gerrit Günther, Constanze Curdt, Martin Weinelt

www.helmholtz-metadata.de
Motivation

Overarching commonality to make content available to researchers.

- 481 entries for Germany (https://www.re3data.org/search?query=&countries%5B%5D=DEU, 2022)
  - F: DOI (218), hdl (37), URN (22), PURL (9), none (159)
  - A: REST (61), OAI-PMH (58), SOAP (11), SPARQL (6), FTP (27)
  - I: DataCite (92), DC (78) ISO 19115 (34), DDI (31), Custom (18)
  - R: License (huge majority), Provenance/Versioning (169), Quality management (275)

- What is inside?
- How many of these systems may a researcher access?
- How many of these systems are still actively maintained?
  → Repository software: 122 other, 188 unknown
**FAIR Digital Objects**

**International Perspective**

- Lots of standardization and conceptual work ongoing
- Different implementation options under discussion
  - PID-based, Linked Data-based
- Some prototypical/demonstrator-like implementations
- 1st International FDO Conference (26.10. – 28.10.)

**HMC Perspective**

- Evaluate FAIR DOs as potential top-level commonality across all research fields
- Realize PID-based implementation
- Work on filling gaps in existing landscape to realize FAIR DOs for the Helmholtz Association
- Agree on common properties every Helmholtz FAIR DO must follow
- Intensive national and international exchange for global alignment

[What are FAIR Digital Objects?](https://fairdo.org/)

© FAIR Digital Objects
Demystifying FAIR DOs: The basic Ingredients

- **Persistent Identifiers**

  - **Globally persistent, unique identification** of digital content
  - **Established, distributed** PID systems available, e.g., handle.net
  - **Long-term guarantee** for PID resolution (≥ 10 years)
  - **Must support** storing **key-value metadata** at PID service → PID Kernel Information

21.T11981/6ab464ed-978b-4996-876f-f68ea913a308

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Demystifying FAIR DOs: The basic Ingredients

- Persistent Identifiers
- DataTypes

- Definition of types of data (fields, structures)
- Based on RDA Recommendation by Lannom et al. [1]
- Described in a machine-readable format
- Stored in Data Type Registry accessible by machines (and humans)
- Globally unique and persistently identified by PIDs

21.T11981/6ab464ed-978b-4996-876f-f68ea913a308

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.T11148/076759916209e5d62bd5</td>
<td></td>
</tr>
<tr>
<td>21.T11148/1c699a5d1b4ad3ba4956</td>
<td></td>
</tr>
<tr>
<td>21.T11148/b8457812905b83046284</td>
<td></td>
</tr>
</tbody>
</table>

PID Kernel Information

[1] https://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458
Demystifying FAIR DOs: The basic Ingredients

- Persistent Identifiers
- DataTypes
- PID Kernel Information Profile

- **Schema for PID Kernel Information** (content of PID Record)
- Based on **RDA Recommendation** by Weigel et al. [1]
- **Strongly relies on PIDs and DataTypes** for describing values
- **Goal**: Provide **machine-actionable metadata** on PID-level for fast decision making

---

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.T11148/1c699a5d1b4ad3ba4956</td>
<td>21.T11148/1a1e620666cb1713acde</td>
</tr>
</tbody>
</table>
| 21.T11148/b8457812905b83046284 | [https://b2share.eudat.eu/api/files/][...]

[1] https://doi.org/10.15497/rda00031
Defining a Helmholtz Kernel Information Profile

- RDA Draft Kernel Information Profile (KIP) defines **15 basic attributes**, mostly administrative information
- Extension of Draft KIP by contextual and relational attributes agreed on between representatives from all research fields
- **Goal**: Increase immediate (scientific) benefit of using FAIR DOs
- **Compatible** to RDA Recommendations
- **Basis for all FAIR DOs** created within the Helmholtz Association
- **Extensible** by additional attributes if required
- **Guidance document available**, publication soon

### Additional Helmholtz KIP Attributes

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>digitalObjectLocation-AccessProtocol</td>
<td>Access information for digitalObjectLocation, e.g., protocol, protocol version, and client</td>
</tr>
<tr>
<td>underEmbargoUntil</td>
<td>Access restrictions probably apply before</td>
</tr>
<tr>
<td>license</td>
<td>Extracted from digitalObjectPolicy</td>
</tr>
<tr>
<td>checksum</td>
<td>Renamed from 'etag' to be more specific</td>
</tr>
<tr>
<td>signature</td>
<td>Cryptographic signature of PID record</td>
</tr>
<tr>
<td>topic</td>
<td>Topic term from vocabulary for additional context</td>
</tr>
<tr>
<td>locationPreview</td>
<td>Optional preview for digitalObjectLocation</td>
</tr>
<tr>
<td>contact</td>
<td>Contact information, e.g., ORCiD or ROR</td>
</tr>
<tr>
<td>hasMetadata</td>
<td>PID pointing to a related FDO containing metadata</td>
</tr>
<tr>
<td>isMetadataFor</td>
<td>Inversion for hasMetadata</td>
</tr>
<tr>
<td>wasGeneratedBy</td>
<td>W3C PROV-DM element to refer to tool/agent used for generating the digital object</td>
</tr>
<tr>
<td>provenanceGraph</td>
<td>Optional PID of full provenance graph</td>
</tr>
</tbody>
</table>
A Demonstrator for the Helmholtz Kernel Information Profile

Architecture

- Implemented by members of CCT4
- Showcase implementation for evaluating applicability for existing repository (Zenodo)
- Blueprint for extension to additional repository platforms
- Basis for constantly growing collection of FAIR DOs

React Web-Frontend

10.5281/zenodo.6561230

User-provided/harvested PID

Provide selected mapper with PID

Provide mapped PID Record attributes

Register + index PID and validated PID Record

Mapper

Typed PID Maker

10.5281/zenodo.6561230

User-provided/harvested PID

Provide selected mapper with PID

Provide mapped PID Record attributes

Register + index PID and validated PID Record

Blueprint for extension to additional repository platforms

Basis for constantly growing collection of FAIR DOs
Demonstrator - Impressions

FAIR Digital Object Demonstrators 2021

Wittenburg, Peter; Anders, Ivonne; Blanchi, Christophe; Buurman, Merret; Goble, Carole; Grieb, Jonas; Hardisty, Alex; Islam, Sharif; Jejkal, Thomas; Kálmán, Tibor; Kirkpatrick, Christine; Lannom, Laurence; Lauer, Thomas; Manepalli, Giridhar; Peters-von Gehlen, Karsten; Pfeil, Andreas; Quick, Robert; van de Sanden, Mark; Schwardmann, Ulrich; Soiland-Reyes, Stian; Stotzka, Rainer; Trautt, Zachary; Van Uytvanck, Dieter; Weiland, Claus; Wieder, Philipp.

This paper gives a summary of implementation activities in the realm of FAIR Digital Objects (FDO). It gives an idea which software components are robust and used for many years, which components are comparatively new and are being tested out in pilot projects and what the challenges are that need to be urgently addressed by the FDO community. After basically only one year of advancing the FDO specifications by the FDO Forum we can recognise an increasing momentum to test and integrate essential FDO components. However, many developments still occur as solisitic engagements that offer a scattered picture. It is widely agreed that it is now time to combine these different pilots to comprehensive testbeds, to identify still existing gaps and to turn some services into components of a convincing and stable infrastructure. This step is urgently needed to convince even more institutions to invest in FDO technology and therefore to increase FAIRness of the evolving global data space.
Demonstrator - Impressions

HMC Kernel Profile Demonstrator

**Dataset:** https://doi.org/10.5281/zenodo.5872645

- KernelInformationProfile
  - 21.T11i4b/tb97/8883245e32c291f7
  - digitalObjectType
    - hard-coded/zenodo_recordType

- digitalObjectLocation
  - https://zenodo.org/api/records/5872645

- digitalObjectLocationAccessProtocol
  - "protocol": "HTTP", "type": "application/json"

**DATE & TIME**

- Date & Time
  - 2022-01-18 14:13

**VERSION NUMBER**

- Semantic Version
  - 1.0.2

- URL
  - https://spdx.org/licenses/CC-BY-4.0.html

**Contact**

- https://orcid.org/0000-0003-3538-0106
- https://orcid.org/0000-0001-7337-3009
- https://orcid.org/0000-0003-2277-5176

**wasQuotedFrom**

- No data

**alternateOf**

- No data

**provenanceGraph**

- [Register FOI]
Conclusions and Outlook

- Agreed on Helmholtz Kernel Information Profile applicable for all Helmholtz FAIR DOs
- Extension of RDA Draft Kernel Information Profile by (mostly optional) contextual attributes
- Implemented first version as demonstrator for mapping digital assets from Zenodo
- Dissemination of results nationally (HMC, NFDIs) and internationally (RDA, EOSC)

- Building a growing collection of FAIR DOs
- Integrate search via Elastic
- Elaborate possibilities for further automation
- Integrate additional repositories and compile guidelines for others
- Identify and implement further applications