HMC | HELMHOLTZ METADATA COLLABORATION

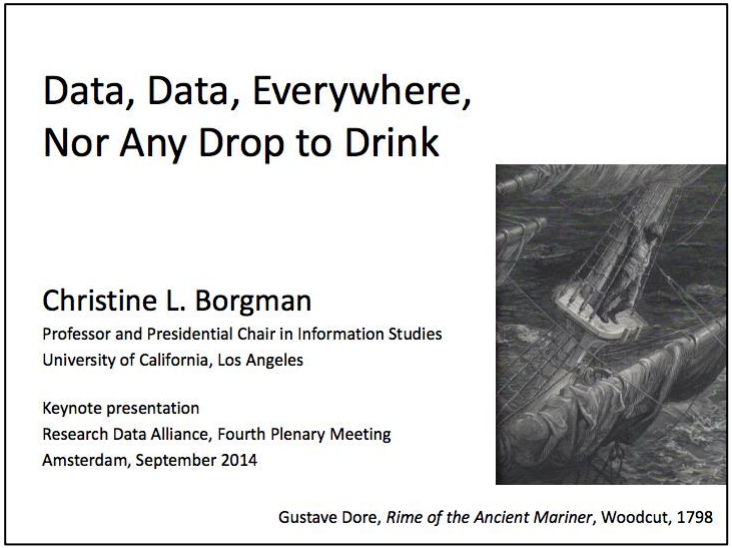
KIT Karlsruhe Institute of Technology · JÜLICH FORSCHUNGSZENTRUM · DLR · dkfz. · HZB Helmholtz Zentrum Berlin · GEOMAR

AST | E&E | ENERGY | HEALTH | INFORMATION | MATTER | FAIR DATA COMMONS | HMC PROJECTS

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

Motivation



- 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

Overarching commonality to make content available to researchers.

shorturl.at/CIJZ7
shorturl.at/lm125

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

Demystifying FAIR DOs: The basic Ingedients

- **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

Key	Value

PID Kernel Information

shorturl.at/ktZ69

Demystifying FAIR DOs: The basic Ingedients

- 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

DataTypes

Key	Value
21.T11148/076759916209e5d62bd5	
21.T11148/1c699a5d1b4ad3ba4956	
21.T11148/b8457812905b83046284	

PID Kernel Information

[1] https://doi.org/10.15497/A5BCD108-ECC4-41BE-91A7-20112FF77458

Demystifying FAIR DOs: The basic Ingedients

- 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

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

Key	Value	
21.T11148/076759916209e5d62bd5	21.T11148/863d938d632b53d62d52	Profile
21.T11148/1c699a5d1b4ad3ba4956	21.T11148/1a1e620666cb1713acde	Type
21.T11148/b8457812905b83046284	https://b2share.eudat.eu/api/files/[...]	Location

DataTypes

PID Kernel Information

[1] https://doi.org/10.15497/rda00031

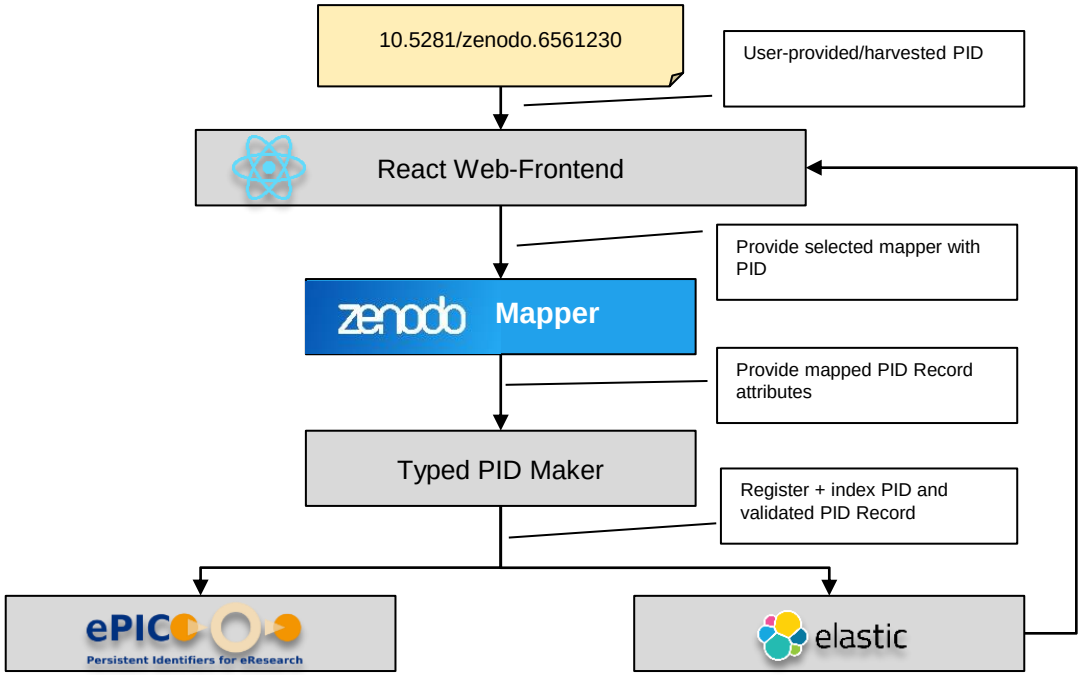
Defining a Helmholtz Kernel Information Profile

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

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**

Demonstrator - Impressions

January 18, 2022

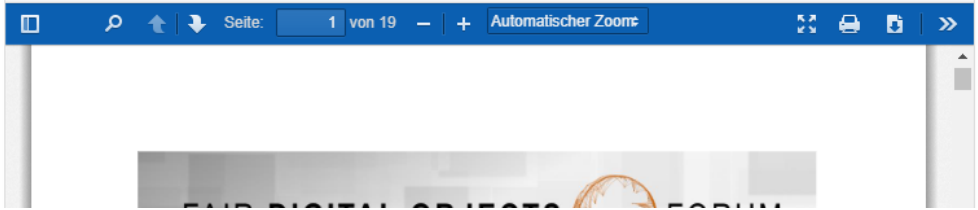
Report Open Access

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 soloistic 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.

Preview



677 views

517 downloads

See more details...

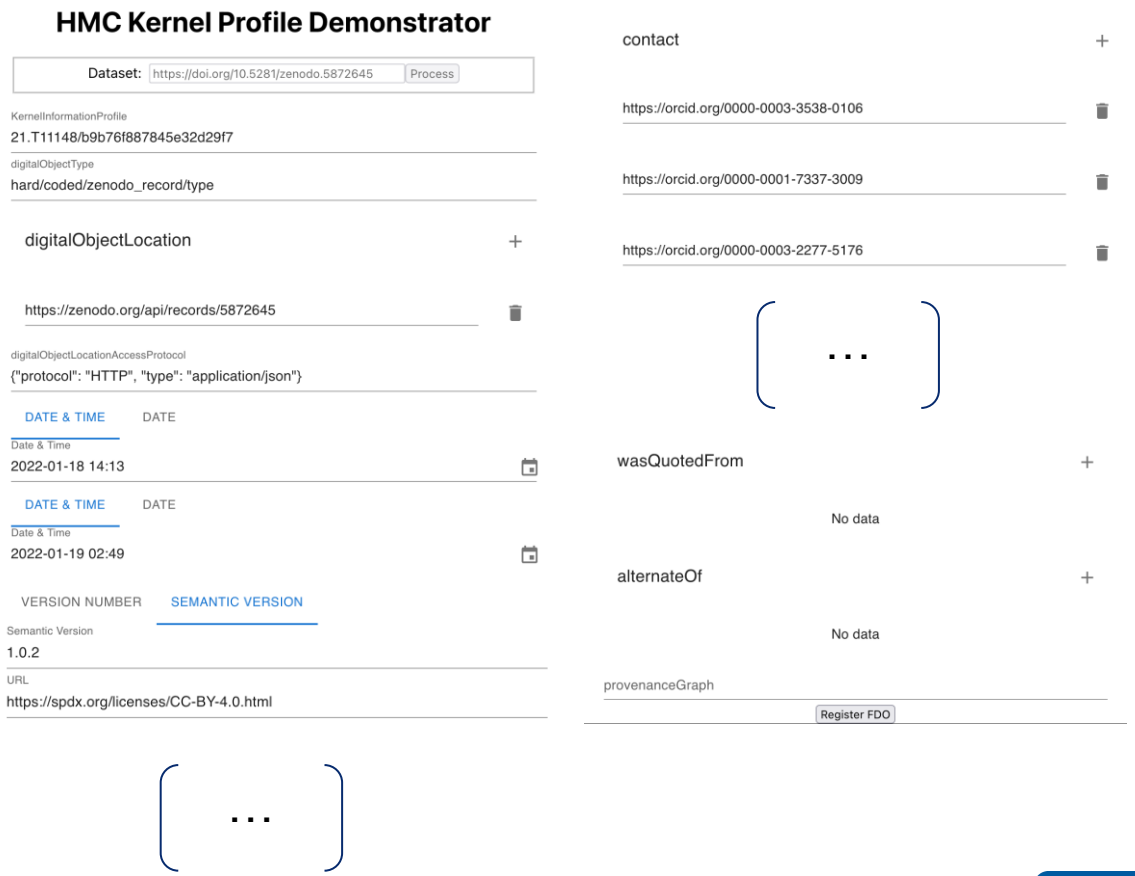
Indexed in OpenAIRE

Publication date:
January 18, 2022

DOI:
DOI 10.5281/zenodo.5872645

Keyword(s):
FAIR, FAIR Digital Objects, Data Management, Data Science, Global Data Space

Demonstrator - Impressions



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

1-31

2-36

2-37

FAIR DO Graph