

Two Examples on How FDO Types can Support Machine and Human Readability

Ulrich Schwarzmänn[‡], Tibor Kálmán[‡]

[‡] Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), Göttingen, Germany

Corresponding author: Ulrich Schwarzmänn (uschar1@gwdg.de)

Abstract

FAIR Digital Objects (FDOs) are typed by a well described set of attributes, where attributes are key value pairs with a key, which refers to a syntactic description of the value. Often the description of the set of attributes is called also profile. The exact description of the attribute keys is obviously crucial for machine actionability on one hand. On the other hand an exact description of attributes can be a way to allow also human readability of the used keys. Furthermore it can often integrate legacy attribute sets that are provided inside repositories for the description of their digital objects.

In the following we show two examples of FDO types with their attributes from different viewpoints. The two examples are: the Persistent Identifiers (PID) for Instruments example and the DARIAH (see <https://de.dariah.eu>) use case. In both cases the Handle system is used for the persistent identifiers, the FDO record is provided by the Handle record of the PID and the attributes can be found here as type-data pairs in the phrasing of the Handle system.

1 Example 1: PID for Instruments

The PID for instrument example goes back to the development of kernel metadata, which is seen as minimally required to reference and describe scientific instruments Stocker et al. 2020. The value space for the attributes here often contains hierarchical objects and can also be lists of attributes.

An example of such an attribute definition is that of a single manufacturer of an instrument that occurs in a list of manufacturers here: <http://dtr-test.pidconsortium.eu/#objects/21.T11148/7adfcc13b3b01de0d875>.

1.1. The Handle Record for a Full PID for Instruments

In this case one uses the references to the attribute definitions as keys for the values, which are often lists or objects. The Handle Record for a full attribute list of a PID for

Instruments can be obtained from the Handle Proxy with <https://hdl.handle.net/21.T11998/0000-001A-3905-1?noredirect>

The structure of this FDO record is defined as a type definition at the ePIC Date Type Registry Schwarzmann 2020 with <http://dtr-test.pidconsortium.eu/objects/21.T11148/17ce618137e697852ea6> . This way we can also refer to this structure definition in a qualified key value pair like TYPE/0.TYPE and then use as keys in the FDO record the names as they are given for keys in this structure. This way an FDO record becomes a human readable form without losing any machine readability. For further details see: <https://hdl.handle.net/21.T11998/0000-001A-3905-8?noredirect>

In both cases the full instrument descriptions are completely stored in the Handle database of the Handle PID service. The PID itself is a metadata object and can be seen as an FDO of its own.

1.2. Type for a PID4Inst based on Attributes

The type for such FDOs is given via proxy <https://hdl.handle.net/21.T11148/17ce618137e697852ea6> in the ePIC DTR

1.3. PID4Inst in a Repository

Another option is to store the metadata objects of instrument descriptions in repositories. In this case a schema is needed to describe the metadata elements that are needed for the description. The existing attribute definitions could be bundled here into a single complex definition, which is syntactically almost identical to the type definition for instruments.

From such a complex definition one could derive a schema for the repository entries. In this case the schema was directly derived from the type, which is conceptually different from attribute definitions, but syntactically similar and therefore exploitable by the same services. The result of the schema derivation can then be directly fed into the ingest module of repositories, in the following figure for example into the CORDRA schema module for the definition of attribute types: <https://hdl.handle.net/21.T11148/c2c8c452912d57a44117>

An example of such a PID for instrument object in a repository is given at <https://vm11.pid.gwdg.de:8445/objects/21.11145/8fefa88dea40956037ec>

2. Example 2: The DARIAH Use Case

This example evolved in the humanities in the DARIAH project about five years ago with the DARIAH repository (Kálmán et al. 2016, DARIAH-DE 2016). The Handle record structure was created far before FDO records have been discussed. It uses key value pairs with human readable keys as the type and provides relatively atomic values. For humans the key here is a description for the value space that can be expected: <https://hdl.handle.net/21.11113/0000-000B-CA4C-D?noredirect>

The use of human readable keys does however not match the goal of machine readability of this description. Additionally it has the risk of uncertainty and ambiguity.

2.1. Attribute Definitions

In order to make these attributes machine readable, attribute definitions for the allowed value spaces were needed and can be found in the ePIC data type registries. The following basic information type for an email address can be used as the reference key for the value space given for the 'RESPONSIBLE' type above for instance: <https://dtr-test.pidconsortium.eu/#objects/21.T11148/e117a4a29bfd07438c1e>

Attribute definitions for all attributes used in the DARIAH example are given in the ePIC data type registrie. This way one is able to define a type for the DARIAH Handle records.

2.2. An FDO Type of Legacy Repository Records

Such a type definition is given at: <https://dtr-test.pidconsortium.eu/#objects/21.T11148/f1eea855587d8b1f66da>

If this type is the known type of all objects in the DARIAH repository, the references to the keys are named very similar the human readable form of the Handle record.

Usually and as we have seen in the previous PID4Inst example the type of the FDO would be another attribute of the FDO. This would require an adaption of the attributes of all digital objects of the DARIAH repository. But since all digital objects of the DARIAH repository follow the same profile and all its digital objects have the same PID prefix, it would be sufficient to implement this additional attribute at the prefix level. Together with a rule that attributes on a lower level dominate attributes on a higher level, this additional prefix attribute makes FDOs out of legacy digital objects that have been defined a long time ago.

3. Presentation

In the presentation we will describe based on the two examples above how machine and human readability can be supported at the same time by FDO types and discuss the implementation of hierarchical type definitions as the basic infrastructure for FAIRness of data in more detail.

Presenting author

Tibor Kálmán

Presented at

First International Conference on FAIR Digital Objects 2022, presentation

Conflicts of interest

References

- DARIAH-DE (Ed.) (2016) The DARIAH-DE Repository. <https://de.dariah.eu/en/repository>. Accessed on: 2022-7-08.
- Kálmán T, Kong X, Schwarzmán U (2016) Die digitale Forschungsinfrastruktur DARIAH-DE: Angebotspalette für die Geistes- und Kulturwissenschaften. Bibliothek Forschung und Praxis[In German]. <https://doi.org/10.1515/bfp-2016-0041>
- Schwarzmán U (2020) The ePIC PID Information Type Registry. Göttingen Research Online / Data <https://doi.org/10.25625/9dnrsj>
- Stocker M, Darroch L, Krahl R, Habermann T, Devaraju A, Schwarzmán U, D'Onofrio C, Häggström I (2020) Persistent Identification of Instruments. Data Science Journal 19 <https://doi.org/10.5334/dsj-2020-018>