



# FDO – Kernel Attributes & Metadata Version 2.0

FDO Forum PR 17 October 2022

## Current and previous versions:

Current Version PR-KernelAttributes-2.0

<https://docs.google.com/document/d/1OF49wTNVuv-6OXINerhBTqVtHyc7jutTaUHjn6BZCs0/edit?usp=sharing>

Previous Version: WD-FDO-KernelAttributesAndMetadata-1.0-20220822;

<https://docs.google.com/document/d/1OF49wTNVuv-6OXINerhBTqVtHyc7jutTaUHjn6BZCs0/edit?usp=sharing>

## Editors:

Daan Broeder, Peter Wittenburg, Ivonne Anders, Karsten Peters-von Gehlen

## Abstract

Since the start of the FDO concept development there has been a discussion on what attributes are needed for proper working FDOs in different contexts and what would be the difference and overlaps with object metadata such as currently used in research data-management. This document draws on previous discussions wrt. attributes stored with Handle type PIDs to facilitate easy data-management and discussions to derive a core-metadata schema for interdisciplinary work.

## Status of this document

This version WD1.0 is the result of the TSIG/BIG commenting phase on the internal document versions<sup>1</sup>. All comments could be resolved so far.

---

1

[https://docs.google.com/document/d/1wFPxdjtVBpY-hFzK\\_iRJea-f8DjTRS-Jtk886r9T9DM/edit#heading=h.9xou53xwi3fo](https://docs.google.com/document/d/1wFPxdjtVBpY-hFzK_iRJea-f8DjTRS-Jtk886r9T9DM/edit#heading=h.9xou53xwi3fo)

The document “PubCom-PR-PIDProfileAttributes-2.0”<sup>2</sup> makes statements about the selection of kernel attributes that should be used in PID records. It states that to achieve machine actionability

- all attributes being used must be registered and defined (incl. those added by communities)
- three categories of attributes are permitted: FDO Mandatory, FDO Optional and Attributes from communities which are always optional<sup>3</sup>.

The first two categories are the result of an FDO process that needs to be established by the BIG group. In this note we would like to elaborate on these three categories. We will refer to the RDA Kernel Type document which also makes suggestions based on the participation of several user communities. This document can also be seen as a commented endorsement of the RDA Kernel Type document to adopt it for the FDO purposes.

## Acknowledgments

This document is based on the work within the GEDE initiative and uptake discussions in the FDO TSIG WG. We thank all colleagues who contributed to this document.

## Contents

<b>Abstract</b>	<b>1</b>
<b>Status of this document</b>	<b>1</b>
<b>Acknowledgments</b>	<b>2</b>
<b>Contents</b>	<b>2</b>
<b>1. Mandatory Attributes</b>	<b>3</b>
<b>2. Recommended Attributes</b>	<b>3</b>
<b>3. Optional FDO Attributes</b>	<b>4</b>
<b>4. FDO Technical Management Responsibilities Information</b>	<b>5</b>
<b>5. Community Provided Attributes</b>	<b>6</b>
<b>6. Community Provided Metadata</b>	<b>6</b>
<b>Appendix A: RDA Kernel Types</b>	<b>6</b>
<b>Changes from previous versions</b>	<b>8</b>

---

<sup>2</sup> <https://docs.google.com/document/d/16zxfnC47vZM4g-dyJoI9XsQMQQHnp11Ajr55F6y6Ww/edit>

<sup>3</sup> Is this correct or would we also allow mandatory attributes defined by communities.

## 1. Mandatory Attributes

The document “PubCom-PR-PIDProfileAttributes-2.0”<sup>4</sup> makes statements about the selection of kernel attributes that should be used in PID records. It states that to achieve machine actionability

- all attributes being used must be registered and defined (incl. those added by communities)
- three categories of attributes are permitted: FDO Mandatory, FDO Optional and Attributes from communities which are always optional<sup>5</sup>.

*Note: it is suggested restricting the mandatory attributes to these two: profile and type. The profile determines the other attributes and this gives the necessary flexibility. The problem is that if we extend the mandatory attributes itself this always implies a restriction of the universality and flexibility of FDOs. Using a profile there is a mechanism to impose everything needed without losing the machine actionability. Some level of interoperability exists, because the profile mechanism and the definition of attributes are (or better can be made) machine actionable like type definitions. But interoperability needs more than machine actionability and by requiring interoperability between all types of FDOs, the set of objects that are FDOs is unnecessarily restricted. For these two mandatory attributes there should be no debate about extending or restricting them anymore and there is no governance necessary. They belong with the FDO concept itself.*

## 2. Recommended Attributes

Recommended Attributes are thought necessary for operating and using FDO's in practice.

- **RDA dateCreated:** the creation of the FDO
- **FDOF ScientificDomain:** This attribute is an indicator of the scientific domain the FDO refers to. This ensures compliance with the FAIR principles, which are per definition applicable at the domain level. This attribute is required since different mandatory attributes may be required at the domain-level.
- **FDOF PersistencyPolicy:** this attribute indicates what the intention of its creator is with respect to its life-time/maintenance, and value domain is a vocabulary with {UNKNOWN, NONE, ##Years} (note: seems only partly covered by RDA digitalObjectPolicy. This value will be set to “UNKNOWN” at creation time.

---

<sup>4</sup> <https://docs.google.com/document/d/16zxfnC47vZM4g-dyJoI9XsQMQQHnp11Ajyr55F6y6Ww/edit>

<sup>5</sup> Is this correct or would we also allow mandatory attributes defined by communities.

- **FDOF ResponsibleOrganisation:** note that this can be another organisation than the PID issuer. the value is taken from the ROR registry value domain (or other with namespace id)

In the case of an FDO Configuration type that includes a bit-sequence the following attributes are mandatory which will be represented in the PID profile:

- **RDA digitalObjectLocation:** In the case that the FDO includes bit-sequences as illustrated for example in FDO configuration type 8 it is mandatory that the PID record contains the location either as an URL or a PID.
- **FDOF digitalObjectType:** The included bit-sequence can be of different type (metadata, time series, etc.). This type needs to be registered in a type registry.
- **RDA digitalObjectPolicy:** Pointer to a policy object which documents changes to the object or its Kernel Information record, including particularly object access and modification policies. A caller should be able to determine the expected future changes to the object from the policy, which are based on managed processes the object owner maintains.
- **FDOF digitalObjectMutability:** This attribute indicates whether the included bit-sequence is mutable or immutable.
- **RDA dateModified:** In the case of mutable bit-sequences of the FDO the last date/time of object modification.
- **RDA checksum:** Checksum of object contents. Checksum format is determined via the attribute type referenced in a Kernel Information record.

### 3. Optional FDO Attributes

To increase interoperability the FDO Forum should define and register a set of attributes that will be frequently used across sectors/disciplines. PID creators should use these attributes as much as possible.

- **RDA version:** If tracked, a version for the object, which must follow a total order. Mandatory for all objects with at least one predecessor version.
- **Prov-DM wasDerivedFrom:** Transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity
- **Prov-DM specializationOf:** Entity is a specialisation of another that shares all aspects of the latter, and additionally presents more specific aspects of the same thing as the latter.
- **Prov-DM wasRevisionOf:** A derivation for which the resulting entity is a revised version of some original.
- **Prov-DM hadPrimarySource:** A primary source for a topic refers to something produced by some agent with direct experience and knowledge about the topic, at the time of the topic's study, without benefit from hindsight.
- **Prov-DM wasQuotedFrom:** Used for the repetition of (some or all of) an entity, such as text or image, by someone who may or may not be its original author.

- **Prov-DM alternateOf:** Entities present aspects of the same thing. These aspects may be the same or different, and the alternate entities may or may not overlap in time.
- **FDOF rightsRecord:** This is a pointer to a possible record in a database that includes access permissions implemented as URL or PID.
- **FDOF licenceConditions:** This is a pointer to a possible record in a database that includes formal specifications about licences such as CC-x implemented as URL or PID.
- **FDOF transactionRecord:** This is a pointer to a possible record in a secure database that includes contractual information implemented as URL or PID.
- **FDOF operationInfo:** Some communities want to include a payload information such as a thumbnail image in the case of DiSSCo's Digital Enhanced Specimen FDO.

In the near future there will be a proper process defined and implemented for deciding and maintaining the lists of mandatory, recommended and optional attributes.

## 4. FDO Technical Management Responsibilities Information

FDOs form part of the technical data infrastructure that needs to be managed by technical infrastructure stakeholder organisations. A distinction should be made between organizations that are responsible for the technical infrastructure and the FDO content, ie. research data in the bitstream or community metadata. From the FDO life-cycle {eg. creation, use, update, retirement} there is a need to establish responsibility and identity of the following roles:

- **FDO Creator** - organisation responsible for creating the FDO (and implicitly issuing the FDO PID), stating RDA DateCreated and XXX:ResponsibleOrganisation, attributes. (optional attribute )
- **FDO Responsible Organisation** - after creation, the same or another organisation will be responsible for further management of the FDO. The Responsible Organisation equals the FDO Creator if available by default (mandatory attribute)
- **FDO auditor** - the auditor organisation is responsible for checking the technical consistency of the FDO. This role will probably be taken also by the FDO Responsible Organisation and only needed from a separate entity in the case of important archive & repository data that need certified availability and access procedures. (optional attribute)

Valid organisations should be registered in recognised organisation registries, the value domain of the attributes includes the registry namespace specification eg. r3data as registry and r3data:r3d100013084 as an identifier for SurfSARA data repository.

The relation between technical FDO management and content responsibilities should be mandatory wrt the following information:

- the Responsible Organisation will initially provide the information (ie. attributes) that indicate the organisation/person responsible for content management / content metadata. After that initial action, the community organisation/person can autonomously

hand over responsibility to another community organisation/person and update the attribute.

- the FDOF PersistencyPolicy is updated by the community organisation/person.

## 5. Community Provided Attributes

Many communities have well defined metadata schemas that contain information that matches the FDO mandatory/optional attributes. Such communities will want to make use of attributes they defined and registered properly. The BIG group of the FDO Forum needs to establish a process to register accepted type registries and a way to integrate the references to such defined types.

## 6. Community Provided Metadata

In many cases community-provided metadata includes many references or information that is relevant at FDO management level. As indicated above, FDO Forum will specify mandatory PID attributes, i.e. the FDO Forum requires that the corresponding attributes need to be extracted from available sources and filled in into the PID record.

FDO managing organisations are strongly recommended to synchronise the FDO attributes from the metadata at creation time and later, whenever metadata is updated.

## Changes from previous versions

This version 1.0 has the status of a WD to be discussed in the FDO Forum. The following changes were made from the TSIG Internal version:

- The FDO Document Template was used.
- Document changes and version history modified

The change history of the TSIG internal document is indicated in this table:

Version			
WD 1.0	Daan Broeder	Aug 28 2022	Deep internal discussions in TSIG and BIG lead to this version.
PR 2.0	Editors	17 October 2022	No further comments were made

## Appendix A: RDA Kernel Types

The RDA recommendation uses the cardinality to indicate mandatory from optional FDO types.

	Property identifier	Content format	Cardi-nality	Explanation
1	PID	Handle	1..n	Global identifier for the object; external to the PID Kernel Information
2	KernelInformation Profile	Handle	1	Handle to the Kernel Information type profile; serves as pointer to profile in DTR. Address of DTR federation expected to be global (common) knowledge.
3	digitalObjectType	Handle	1	Handle points to type definition in DTR for this type of object. Distinguishing metadata from data objects is a client decision within a particular usage context, which may to some extent rely on the digitalObjectType value provided.
4	digitalObjectLocati on	URL	1..n	Pointer to the content object location (pointer to the DO). This may be in addition to a pointer to a human-readable landing page for the object.
5	digitalObjectPolicy	Handle	1	Pointer to a policy object which documents changes to the object or its Kernel Information record, including particularly object access and modification policies. A caller should be able to determine the expected future changes to the object from the policy, which are based on managed processes the object owner maintains.
6	etag	Hex string	1	Checksum of object contents. Checksum format determined via attribute type referenced in a Kernel Information record.

7	dateModified	ISO 8601 Date	0..1	Last date/time of object modification. Mandatory if applicable.
8	dateCreated	ISO 8601 Date	1	Date/time of object creation
9	version	String	0..1	If tracked, a version for the object, which must follow a total order. Mandatory for all objects with at least one predecessor version.
10	wasDerivedFrom	Handle	0..n	<a href="#">PROV-DM</a> : Transformation of an entity into another, an update of an entity resulting in a new one, or the construction of a new entity based on a pre-existing entity.
11	specializationOf	Handle	0..n	<a href="#">PROV-DM</a> : Entity is a specialization of another that shares all aspects of the latter, and additionally presents more specific aspects of the same thing as the latter.
12	wasRevisionOf	Handle	0..n	<a href="#">PROV-DM</a> : A derivation for which the resulting entity is a revised version of some original.
13	hadPrimarySource	Handle	0..n	<a href="#">PROV-DM</a> : A primary source for a topic refers to something produced by some agent with direct experience and knowledge about the topic, at the time of the topic's study, without benefit from hindsight.
14	wasQuotedFrom	Handle	0..n	<a href="#">PROV-DM</a> : Used for the repeat of (some or all of) an entity, such as text or image, by someone who may or may not be its original author.
15	alternateOf	Handle	0..n	<a href="#">PROV-DM</a> : Entities present aspects of the same thing. These aspects may be the same or different, and the alternate entities may or may not overlap in time.



