FAIR Digital Object Framework

Version 1.02, November 2019

FDOF Technical Implementation Guideline

*"We need a set of principles that are sufficiently specific to be useful but sufficiently abstract to exclude specific software stacks, i.e., a document that will still make sense and still be useful ten years from now."*

This document includes some generic guidelines to be met (chapter 2), a normative part defining the FAIR Digital Object Framework (FDOF) at an abstract level which will develop over time (chapter 3) and a glossary of terms (chapter 4). Related documents such as implementation examples can be found at the Github site.

## Change History

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Intention | Actors |
| Version 1.0 | October 2019 | prepared for the consensus meetings in Washington and Paris in October 2019 | created by Luiz Bonino and Peter Wittenburg |
| Version 1.01 | 17.11. 2019 | created after the consensus meeting in Paris at 28/29.10.2019 | changes by Luiz Bonino and Peter Wittenburg |
| Version 1.02 | 22.11. 2019 | created after various comments | changes by Peter Wittenburg |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Changes from V1.0 to V1.01**

* Restructuring the Document and improving some formulations.
* Leaving out concretization footnotes from the normative part.
* Leaving out footnotes about matters explained in the glossary.
* Making more statements about metadata to indicate their importance.
* **Changes from V1.01 to V1.02**
* The illustration examples of possible implementations were separated from the FDOF core document

## Generic Guidelines

Some overall guidelines need to be met by the FAIR DO Framework (FDOF).

**G1**: Show a path for infrastructure investments for **many decades.**

**G2:** Demonstrate **trustworthiness** to researchers and developers to become engaged.

**G3**: Offer compliance with the **FAIR principles** being turned into indicators of FAIRness by an RDA Working Group (<https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>).

**G4**: Support **machine actionability** which includes referential integrity, which states that all references need to be valid without temporal limitation, and explicitness of semantic relationships.

**G5**: Support the **abstraction principle**, i.e. abstract away from details that are not needed at a specific layer. At the management layer there is no difference to be made between data, metadata, software, semantic assertions, etc.

**G6**: Support **stable binding** between all informational entities that are required for machines to act.

**G7**: Support **encapsulation** which means that operations can be associated with types of FDOs.

**G8**: Support **technology independence** allowing implementations using different technologies

## Requirements for FDOF

The requirements for FDOF describe rules that need to be met by any implementation of the FDO concept. These requirements will develop dependent on the insights.

**FDOF1**: A PID, standing for a globally unique, persistent and resolvable identifier, is assumed to be the basis of the Internet of FAIR Data and Services.

**FDOF2**: A PID is resolved to a structured record with attributes which are semantically defined within a type ontology which can have different forms.

**FDOF3**: The structured record includes at least a reference to the locations where the bit-sequences encoding the content of a FAIR-DO (FDO) can be accessed, a PID pointing to the metadata FDO(s) describing properties of it and the DO's type.

**FDOF4**: The structured record can include other attributes that are important to characterize specific types of FDO or that are required by applications. It is required that these attributes are registered in a type registry.

**FDOF5**: Each FDO identified by a PID can be accessed or operated on using an interface protocol by specifying the PID of a registered operation and the PID of the access point.

**FDOF6**: This protocol offers the typical CRUD operations on FDOs and a possibility to use extended operations.

**FDOF7**: The relations between FDO Types and operations are maintained in a type ontology.

**FDOF8**: Metadata descriptions being FDOs and describing the properties of the FDO are made available as semantic assertions enabling machines to act.

**FDOF9**: Metadata assertions can be of different types such as descriptive, deep scientific, provenance, system, access permissions, transactions, etc.

**FDOF10**: Metadata schemas are maintained by communities of practice. FDOF requires that such metadata are FAIR.

**FDOF11**: A collection of FDOs is an FDO and semantic assertions are to be used to describe their construction, i.e. the relationships of their constituents.

**FDOF12**: The "Deletion" of a FDO leads to standardised and thus machine interpretable tombstone notes in the metadata and PID records, i.e. PIDs and metadata should not be deleted.

## Appendix: Glossary

A short glossary with explanations about crucial terms such as "repository", "encapsulation" etc. will help in clarifications, since some terms may be interpreted differently by the participants.

|  |  |
| --- | --- |
| Term | Explanation |
| abstraction | Abstraction is a conceptual process where general rules and concepts are derived from the usage and classification of specific examples. literal signifiers, first principles or other methods (Wikipedia) |
| binding | With binding we mean the possibility for humans and machines to find other relevant entities of a DO when being exposed to another, i.e. when an actor receives a PID of a DO it must find the PID of the corresponding metadata DO and the access rights information, since otherwise interpretation and access is impossible |
| collection | A collection is a complex DO consisting of other DOs, that have a PID and metadata. |
| CRUD operations | These are the usual primary type of operations such as create, read/retrieve, update and delete |
| encapsulation | Encapsulation is known from abstract data types and oo programming where internals of data objects are hidden to the user and where the user can only influence the internal state by using defined methods |
| **Note**: in the FDO case DO types can be associated with registered operations that can be used to operate on DO's content |
| machine actionability | With machine actionability the capacity of computational systems is meant to find, access, interoperate and reuse data and services without human intervention (GOFAIR) |
| metadata | Metadata descriptions of DOs are sets of assertions describing properties of DOs content which are required for finding, accessing, interpreting and reusing, these assertions can cover a wide range such as descriptive to support finding, deep scientific to support science, systemic to support management, rights to prevent unauthorized access, etc. |
| **Note**: Yet the domain of metadata is not structured very well, i.e. terminology is not well-defined. |
| **Note**: Basic interoperability assumptions are that the schemas are registered and the concepts defined and registered. |
| repository | **DO View**: from the perspective of Digital Objects repositories are nothing else than a complex DO associated with a PID, metadata of different kinds and functions to offer DOs |
| **Common View**: from the most common point of view repositories are entities that host data, metadata etc., apply trustworthy management procedures, offer a search and access interface, have a team of experts taking care and have a sustainability plan |
| **Note**: repositories can be associated with research organisations, communities or projects, they can be small or big in terms of the collections they hold. |
| type | "Type" is an attribute of digital objects which tells computational actors how the content of the DO needs to be parsed, i.e. it defines the operations that can be done on the data, the meaning of the data, and the way values of that type can be interpreted |
| **Note**: A MIME type is a standard that indicates the nature and format of a document, file, or assortment of bytes, i.e. it is a restricted concept of type. |
| **Note**: A type of a DO implies a summary of otherwise complex metadata assertions describing the format, encoding etc. of a content. |