

# Towards a Conceptual Model for the FAIR Digital Object Framework

Luiz Olavo BONINO DA SILVA SANTOS<sup>a,b,1</sup>, Tiago Prince SALES<sup>a</sup>,  
Claudenir M. FONSECA<sup>a</sup> and Giancarlo GUIZZARDI<sup>a</sup>

<sup>a</sup>*Semantics, Cybersecurity & Services (SCS), University of Twente, The Netherlands*

<sup>b</sup>*Biosemantics group, Leiden University Medical Center, The Netherlands*

**Abstract.** The FAIR principles define a number of expected behaviours for the data and services ecosystem with the goal of improving the findability, accessibility, interoperability, and reusability of digital objects. A key aspiration of the principles is that they would lead to a scenario where autonomous computational agents are capable of performing a “self-guided exploration of the global data ecosystem,” and act properly with the encountered variety of types, formats, access mechanisms and protocols. The lack of consistent support for some of these expected behaviours by current information infrastructures such as the internet and the World Wide Web motivated the emergence, in the last years, of initiatives such as the FAIR Digital Object (FDO) movement. The FDO aims at defining an infrastructure where digital objects can be exposed and explored according to the FAIR principles. In this paper, we report the current status of the work towards an ontology-driven conceptual model for FAIR Digital Objects. The conceptual model covers aspects of digital objects that are relevant to the FAIR principles such as the distinction between metadata and the digital object it describes, the classification of digital objects in terms of both their informational value and their computational representation format, and the relation between different types of FAIR Digital Objects.

**Keywords.** FAIR, ontology-driven conceptual modelling, FAIR Digital Objects, interoperability

## 1. Introduction

The evolution of the informatics infrastructure has happened so far in incremental and complementary steps. Whenever the challenges presented in one step are addressed, a new set of possibilities emerge and, with them, new challenges. From the challenges of interconnecting computers in a network to interconnecting different networks (the internet) and interlinking networked resources (the World Wide Web), one new layer of interoperability facilitation leverages the previous one.

In the digital realm, we constantly interact with different types of entities (or objects). The totality of bits in a computer’s storage medium is divided into different units, representing different types of objects and distinguished based on their different informational value. For instance, the bits sequence representing a text document, which con-

---

<sup>1</sup>Corresponding Author: Luiz Olavo Bonino da Silva Santos, l.o.boninodasilvasantos@utwente.nl.

tains an essay, constitutes a different informational unit than another bits sequence representing the text editor software.

The FAIR principles [1] brought a vision of a global ecosystem of dynamically interoperable objects, such as data, services and computing capacity. This vision includes computational agents that are “*capable of autonomously and appropriately acting when faced with the wide range of types, format, and access mechanisms/protocols that will be encountered in their self-guided exploration of the global ecosystem*”.

This envisioned scenario requires that the exploring computational agents are provided with enough actionable information about the encountered objects so that they can automate as much as possible the object’s discovery, access, interoperation and reuse. The provision of information about digital objects entails not only that the information is made available to the agents, but it is also required that the infrastructure, participating systems and services expose a number of behaviours for the interactions involved in this provision to be supported. Therefore, to fully realise that vision, we need an infrastructure capable of supporting the manipulation of digital objects according to the requirements defined by the FAIR principles as well as definitions of the expected behaviours of the participating systems, applications and services.

These challenges to applying the FAIR principles in the current digital communication infrastructure are the main motivators of the work on the FAIR Digital Object Framework (FDOF). The FDOF is a framework designed to support the presentation of informational objects in a digital environment. As the name suggests, the FDOF is inspired by the notion of Digital Objects introduced in [2] and extended to comply with the additional requirements derived from the FAIR principles.

The FDOF aims to define features to provide infrastructural support for the FAIR principles. This set of features added to the underlying communication infrastructure is concentrated in requirements and implementation guidelines for a predictable identifier resolution behaviour and a set of basic and extended operations for FAIR Digital Objects, a retrieval mechanism for object metadata from the object’s identifier and an object typing system.

To define the entities and relations included in the framework we have defined an ontology-based conceptual model named FAIR Digital Object Framework Conceptual Model (FDOF-CM), which is the focus of this paper. The remainder of the paper is organised as follows: Section 2 introduces the FAIR Digital Object Framework and discusses the ontological foundations used in our proposed conceptual model; Section 3 presents the FDOF Conceptual Model detailing its main entities and relations; Section 4 evaluates the FDOF-CM by deriving an OWL ontology which is validated in a set of relevant use cases; Section 5 discusses our related work; and Section 6 presents final considerations and future work.

## **2. Research Baseline**

The focus of this paper is the conceptual model defining and describing the entities and relations present in the FAIR Digital Object Framework. Below we provide a research baseline of the work on the FDOF and the ontological notions used in the FDOF-CM.

## 2.1. FAIR Digital Object Framework

In an increasingly complex digital environment automation becomes a necessity and to support increasing levels of automation, proper identification, and classification of different digital objects become relevant as well as the qualification of the relations between them. This brings additional requirements to the digital ecosystem, which are not properly covered by the Internet or the World Wide Web. The latter, for instance, is based on a network of digital objects linked through unqualified links, the web links, i.e., the links between web resources are not semantically defined. The Semantic Web<sup>2</sup> tackles some of these issues such as providing qualified relations between objects and support for semantic annotations. The FDOF uses Semantic Web techniques to represent the FDOF-CM when used for semantic annotations of data and metadata. The parts of the framework dealing with the identifier resolution protocol, which includes a mechanism to retrieve the object's metadata from its identifier will be described in an upcoming paper.

In the seminal paper where the FAIR principles have been introduced [1], the authors list enabling capabilities that autonomously-acting computational data exploring agents acquire from information provided by the underlying infrastructure when facing newly discovered digital objects. These capabilities are: “(1) *identify the type of object (with respect to both structure and intent)*, (2) *determine if it is useful within the context of the agent's current task by interrogating metadata and/or data elements*, (3) *determine if it is usable, with respect to license, consent, or other accessibility or use constraints*, and (4) *take appropriate action, in much the same manner that a human would.*”

These required capabilities can be converted into the following questions the framework should support and enable both computational agents and humans to answer: (Q1) What type of object does this encountered identifier identify? (Q2) How can I get more information (e.g., metadata) about this object? (Q3) How can this object be manipulated, by whom, and under which conditions?

Given their different nature, digital objects can be manipulated in different ways. The manipulation of digital objects occurs in a combination of computational agents and the underlying technological infrastructure. The interactions happen considering at least two aspects of the digital objects, their *representation or materialisation format* and their *informational content*. The former distinguishes ways of manipulating different materialisation formats such as an executable program, a PDF file, or a video file. While the executable program can be run on a given operating system, we can “open” the PDF file for reading or editing using specific supporting software, and we can play the video file to watch the animated sequence using a video player software. In other words, the possibility of interacting with a given digital object depends, in the first moment, on the ability of the computational agent or technological infrastructure to handle the object's encoding format.

The informational aspect of the digital objects is dealt with by inspecting the content of the bit sequences considering their different informational roles, e.g., does a given bit sequence represent a structured dataset, a software code, a natural language text document, or images? Many materialisation formats are directly related to specific types of informational content, e.g., the MP4 format is associated with video sequences and the TXT format is associated with textual content. But in other cases, this relation is not

---

<sup>2</sup><https://www.w3.org/standards/semanticweb/>

so clear. For instance, we can materialise digital images, text, and structured data using PDF files. In these cases, knowing the materialisation format is not enough to clarify the informational nature of the digital object. If we want to increase the autonomy of computational agents in their handling of digital objects, we should also provide information about each object's informational value besides its materialisation format.

In order to support these questions and challenges the FDOF-CM should rely on a set of clear and unambiguous definitions of the involved concepts and their relations. The proposed conceptual model discussed in section 3 aims at providing these clear and unambiguous conceptualizations related to the FDOF. Our focus here is, therefore, the presentation, description, and discussion of the FDOF conceptual model. For a more in-depth explanation of the other part of the framework beyond the conceptual model, its motivation, and historical evolution, the FDOF documentation [3] should be consulted.

## 2.2. Ontological Foundations

Conceptual models are artefacts developed to provide semantic clarification of a subject domain. One approach to building conceptual models is to employ foundational ontologies that assist the modeller in consistently representing fundamental aspects of the domain. We refer to this approach as ontology-driven conceptual modelling (ODCM) [4]. We use in this paper an ODCM extension of UML called OntoUML [5,6]. OntoUML enriches UML's class diagram language with constructs (e.g., classes and relations) that reflect the ontological categories of the Unified Foundational Ontology (UFO) [5,7]. Additionally, OntoUML defines *semantically-motivated syntactical constraints* designed to ensure that valid models represent sound UFO ontologies.

OntoUML uses special labels called stereotypes to decorate classes and relations that express certain ontological categories from UFO. The stereotyped classes, hereby referred to as types, reflect two orthogonal classifications criteria from UFO: (i) classifications based on the ontological nature of the type's instances (e.g., objects, individualized properties, abstract values); and (ii) classifications based on modal characteristics of how the type applies to its instances (e.g., necessarily or accidentally). We now summarize the stereotypes later employed in Section 3.

The UFO ontology employs the notion of identity principle to refer to the best-suited criteria to necessarily classify the different sorts of concrete individuals in a domain. For example, a modeller could select the type Organization as the type providing the identity principle of its instances within a taxonomy ranging from Entity to Professional Limited Liability Organization. In OntoUML, these *ultimate sortals* (also known as *kinds*) are represented by a set of stereotypes that identify both identity principle and the ontological nature of their instances. The stereotype «kind» identifies ultimate sortals whose instances are objects in the domain (e.g., Organization, Person, Car), while «relator» identifies ultimate sortals whose instances are individualized relational properties (e.g., Marriage, Employment, Enrollment). The stereotype «mediation» decorates the associations between these types of relational properties and the types of objects on which their instances are existentially dependent (i.e., their *bearers*).

While ultimate sortals uniquely and necessarily classify the entities following the identity principle it provides, other types of concrete individuals either classify individuals following a common principle or classify individuals of a variety of principles. Classes decorated by «subkind» or «role» inherit a principle provided by an ultimate

sortal, where the former necessarily classifies its instances (e.g., Limited Liability Organization, Sports Car), and the latter does so accidentally based on some relational property (e.g., Hired Organization, Rented Car). Classes decorated with «category» or «roleMixin» are similar to subkinds and roles in classifying instances either necessarily (e.g., Agent, Vehicle) or accidentally (e.g., Acting Agent, Damaged Vehicle), but these classify concrete individuals following different identity principles, i.e., instances of multiple kinds.

Types decorated by «referenceStructure» and «referenceRegion» [8] classify abstract individuals representing structured conceptual spaces (e.g., the three-dimensional Colour Space) and individualized regions within these spaces (e.g., the Colour Value *Crimson-H348°-S91%-B86%*) respectively. Types decorated by «type» are higher-order types whose instances are themselves other types that are part of the subject domain [9], thus, allowing modellers to describe them and their (higher-order) properties much in the same way one describes regular individuals. Associations decorated with «instantiation» reify the notion of instantiation in the model connecting these higher-order types to their so-called *base types*, i.e., the super types of their instances, providing information about the characteristics their instances must inherit and forming the powertype pattern. Furthermore, associations decorated with «historicalDependence» indicate historical dependencies of instances of the association's target on instances of the source [10]. Finally, the relation of *redefinition* of association ends from UML has also been employed in Section 3. This allows the modeller to bind association ends connected to specialised types, to the association ends connected to their parents, expressing a refinement of their parents' association with additional constraints.

To facilitate the readability of the ontological natures captured in OntoUML classes' tagged values, we use a colour coding to denote the ontological nature of their instances, with red denoting objects, green denoting relational properties, purple denoting types, and white denoting abstracts.

### 3. The FAIR Digital Object Framework Conceptual Model

The goals of the FAIR Digital Object Framework Conceptual Model are to define the main concepts of the framework and how they relate and to provide the basis for an ontology of these concepts and relations. This ontology should be used to semantically annotate the elements of the framework in metadata records, semantic data and metadata models, or whenever the representation language used in the digital object allows the inclusion of semantic annotations, e.g., RDF, OWL.

To facilitate understanding, we have divided the FDOF-CM into three parts: object identification, basic typing system, and the distinction between metadata and the object it describes.

#### 3.1. FDO Identification

Proper identification of different digital objects is a central point in the FAIR principles. Three principles explicitly mention identifier, namely: “F1. (*meta*)data are assigned a globally unique and persistent identifier”, F3 (“*meta*data clearly include the identifier of the data it describes”) and A1 (“(*meta*)data are retrievable by their identifier using a standardized communication protocol”).

Figure 1 presents a fragment of FDOF-CM where a FAIR Digital Object is defined as a specialisation of Identified Object. However, differently from an Identified Object that is generically identified by an Identifier, FDOs are identified by a specialisation of Identifier that is globally unique, persistent, and resolvable (GUPRI). The refinement from Identified Object to FAIR Digital Object and from Identifier to GUPRI is a consequence of the requirements of the FAIR principles F1 (for the global uniqueness and persistence) and A1 (for being resolvable) that a FAIR Digital Object not only needs to be identified by an identifier but this identifier must guarantee global uniqueness (single reference), persistence (rigid designation) and be resolvable (retrievable from its identifier).

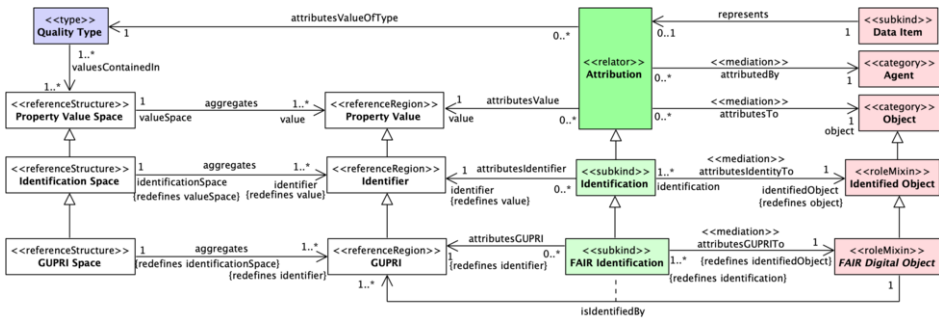


Figure 1. FDOF - Identification and attribution

The model in Figure 1 defines that Identifiers are Property Values within a specialisation of Property Value Space named Identification Space. The Identification Space represents all possible values (identifiers) of that identification system. One example is that the identification space for Unified Resource Identifiers (URIs) is the set of possible values for URIs as defined in [11] following the pattern of a scheme, a colon (:), a single or double forward slash (/ or //), an authority, a path, a query, and a fragment.

Another aspect covered in this FDOF-CM identification fragment is the reification of the Identification relation between Identifiers and Identified Objects, and more specifically the FAIR Identification relation between GUPRIs and FDOs, that are made by Agents. In this way, we can capture the information of who assigned an identifier to a given FDO, adding to the object’s provenance.

### 3.2. FDOF Basic Typing System

Types are taken here to be aggregations of properties that are used to characterise their instances. Commonly, the choice is made to fulfil a particular goal or use. In the realm of digital objects, different typing systems have been used for different purposes, mainly related to creating abstractions for the sets of bits stored in computer memory. For instance, most operating systems use the abstraction of files to organise groups of bits whereas datasets, potentially spanning over several files or as the content of database management systems, are commonly used in data science activities.

On the internet, types of digital objects are often associated with Media Types (formerly known as MIME types) [12]. By looking at the hundreds of types defined by the Internet Assigned Numbers Authority (IANA) as Media Types, one cannot avoid notic-

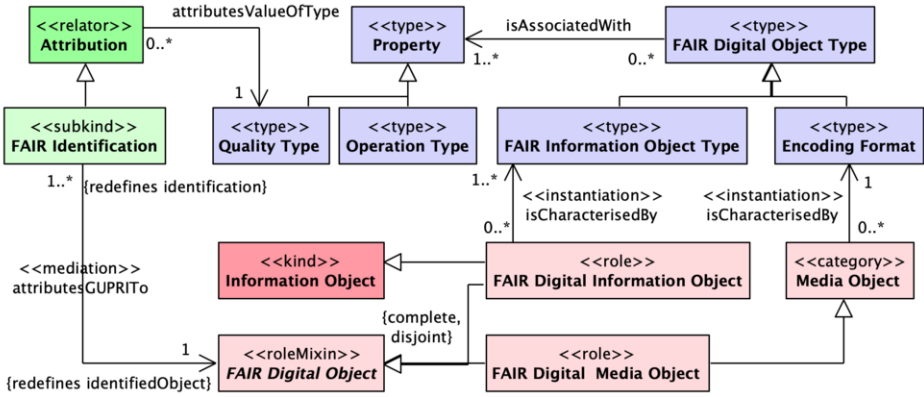


Figure 2. FDOF - FAIR Characterisation

ing some inconsistency in the classification criteria when taking into consideration the nature of the defined entities. In some cases, the media type represents a given file encoding format, e.g., *application/dicom*, *audio/mpeg*, *image/png* or *video/h264*. In other cases, the media type represents not only a generic encoding format but how the content is organised in a particular structure, e.g., the IANA media type *application/alto-directory+json* defines both the encoding format (*JSON*) and a specific structure of key-value pairs following the Application-Layer Traffic Optimization (*ALTO*) protocol.

We consider that this typing system based only on the encoding format is insufficient to support the levels of automation envisioned by the FAIR principles. For instance, when trying to classify a given photo as an FDO, should the type of the FDO represent the image captured in the photo, the encoding format of the photo file or what is depicted in the photo? We argue that we should consider these three aspects as part of the characterisation of a given FDO, namely, (i) the encoding format, (ii) the type of the object regarding its informational content and (iii) the entity(ies) represented by the object. The motivation for the argument is that a computational agent benefits from information in these three levels of abstraction to improve its machine-actionability as required by the FAIR principles. The encoding format helps determine how the digital object can be processed by the computing infrastructure (e.g., which parser or reader should be used). The type of the object considering its informational content helps differentiate, for instance, images from text, video, applications, etc. Finally, the reference to the entities represented by the object provides more information about what the object represents.

Fig. 2 depicts a fragment of the FDOF-CM containing the main classes of the FDOF basic typing system. In this model, a FAIR Digital Object can be either a FAIR Digital Information Object (FDIO) or a FAIR Digital Media Object (FDMO). The FDIO class represents FDOs as units of information, and it inherits the identity criteria provided by Information Object which consists of the objects’ informational contents. Every FDIO is materialized by at least one FDMO. The FDMO is a bit sequence,<sup>3</sup> which (as all Media Objects) encodes characteristics that are proper to the media type (representing here by Encoding Format). Every FDMO also encodes (i.e., materializes) the informational content of at least one FDIO. In the photo example, John’s birthday party photo (an instance

<sup>3</sup>The kind Bit Sequence in Fig. 3 defines a self-evident principle of identity inherited by FDMOs.

of FDIO) is materialised by an object named *JohnBDayPhoto1.jpg*, which is an instance of FDMO encoded as an *image/jpg* (instance of Encoding Format). We could also have the same FDIO materialised by another FDMO instance named *JohnBDayPhoto2.heic* encoded as an *image/heic*. In this example, we have three FDOs involved, one FAIR Digital Information Object, John’s birthday party photo, and two FAIR Digital Media Objects, *JohnBDayPhoto1.jpg* and *JohnBDayPhoto2.heic*.

In general, a given FDO is characterised by one sub-type of FAIR Digital Object Type (FDOT). FDOTs aggregate the set of properties that characterise information objects of that type (see *isAssociatedWith* relation in Fig. 2). For instance, if we have the photo of John’s birthday party, we can classify this object as a FDIO having the information object type *Photo*. This *Photo* FDOT aggregates a number of properties that characterise photos, e.g., the event depicted, the participants, location, date and time, etc. However, as any Object here, this photo can have a number of attributions referring to other Identified Objects (e.g., an *participant* attribution pointing to an Identified Object of type Person, or a *location* attribution pointing to a property value in a geo-coordinate space).

### 3.3. Metadata

One important element of the FAIR principles is the role attributed to metadata as responsible for describing a given object (principle F2, data are described with rich metadata). Many of the FAIR principles indicate that they apply to both metadata and other types of objects (principles F1, F4, A1, I1, I2, I3, R1.1, R1.2 and R1.3 [1]). Because of this, the FDOF-CM defines the FAIR Metadata Record (FMR) as a specialisation of FDIO with the role of describing FDOs, as depicted in Fig. 3.

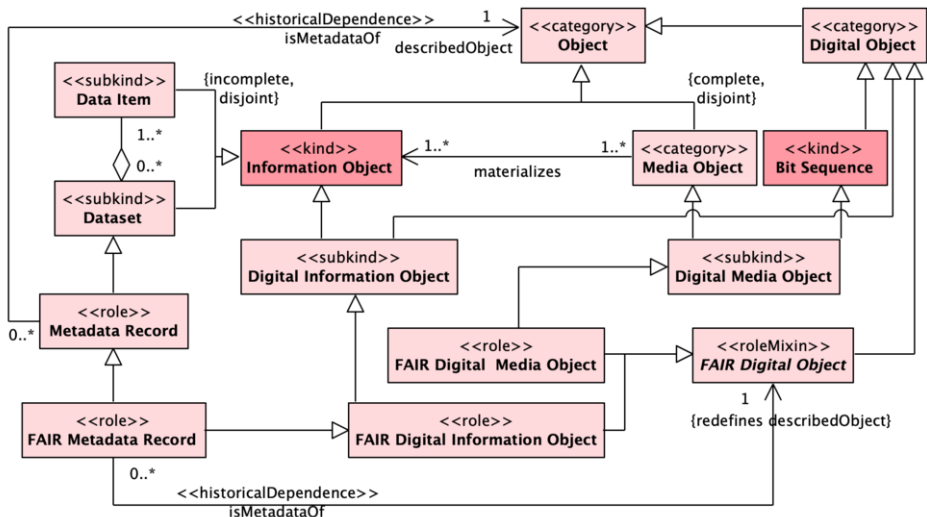


Figure 3. FDOF - Metadata Core

In the model, every FMR is a Dataset, which is an aggregation of Data Items. In this way, a FMR is a sort of dataset information object containing multiple data items that



represent attributions of values to properties describing a particular object. As a specialisation of FDIO, the FMR is characterised by a FAIR Information Object Type. FDOF-CM supports the possibility of defining specialisations of FAIR Information Object Type that associate the sets of properties that can (optional) or should (mandatory) be used to characterise the metadata record. In an implementation, these definitions can be realised in terms of metadata profiles associated with given types of FDOs.

The class FDO is *abstract* (in the UML sense), i.e., cannot be directly instantiated. Instances of FDO must be either instantiate FAIR Digital Information Object or FAIR Digital Media Object. Consequently, both FDIOs and FDMOs should be described by metadata records.

However, the requirement that an FDO must be described by a metadata record and the fact that a FMR is a specialisation of FDO creates the possibility of an infinite loop of metadata records requiring meta-metadata records describing them. To avoid this loop and to better reflect the reality, the FDOF-CM includes the following constraint rule:  $\text{FAIRDigitalObject}(x) \wedge \neg \text{FAIRMetadataRecord}(x) \rightarrow \exists y (\text{FAIRMetadataRecord}(y) \wedge \text{isMetadataOf}(y,x))$ . This rule redefines the zero-to-many cardinality of the *isMetadataOf* relation between FMR and FDO to require at least one metadata record to describe the FDO unless the FDO is an instance of FMR. With this constraint rule, the model still allows metadata records to have their own metadata records but does not mandate it.

As depicted in Fig. 3, we make a distinction between digital and non-digital objects. This supports the possibility of having, for instance, FAIR Metadata Records describing physical objects. However, in this paper, we focus only on digital objects. In fact, in the FDOF-CM we have generic elements such as Identified Object, Information Object, Media Object and Metadata Record. But what makes them FAIR? According to the FDOF, an Identified Object becomes a FDO, regarding identification, when it is identified by a GUPRI and not any other type of identifier, as depicted in Fig. 1. Similarly, a generic Metadata Record becomes a FMR when is assigned with a GUPRI and includes the identifier of the object it describes (per FAIR principle F3) using the relation *isMetadataOf* (Fig. 3). Regarding metadata and characterisation, an object becomes a FDO when it is characterised by its informational value (using FDIO), its materialisation format (using FDMO), and it is described by a FMR.

### 3.4. Ontological Considerations

The FDOF-CM has been designed with the requirement of making it as simple as possible so that it can be understood in a wider context including all different types of web practitioners. However, it incorporates a number of ontological ideas and design patterns.

The mechanism of type characterization in Figs. 1 and 2 embeds the elements of the *Four-Category Ontology* pattern (found in Foundational Ontologies such as UFO, DOLCE, GFO and BFO). Here we have: an Object instantiates an Object Type and is the *bearer of a quality* that *exemplifies* a given Quality Type. Types here are defined as aggregations of Properties (as in *hyperintensional* type definitions). Actual quality individuals are represented by Attributions that connect the bearer (the Object at hand) with a projection of that quality as a Property Value in a given Property Value Space. This extends the classical Four-Category Ontology into the so-called *Ontological Octagon* as proposed by [13].

The mechanism in Fig. 3 that treats FAIR Metadata Records (FMRs) as FDIOs allows for the former to instantiate proper FAIR Digital Object Types and hence be characterized by properties, but also be described by other FMRs, themselves FDIOs that can instantiate FAIR Digital Object Types characterized by Properties and described by FMRs, and so forth. This implements a minimum set of features available in more sophisticated multi-level theories such as proposed in [14].

The relation between Information Objects and Media Objects seem at first to reflect the relation of *generic dependence* between content entities, e.g., *The Aleph* by J.L.Borges (as a particular content with fictional individuals, locations and events) and a particular physical copy of that book standing on my table, i.e., the notions of *Information Content Entity* and *Information Bearing Entity* in the Information Artifact Ontology (IAO) [15]. However, notice that a Digital Media Object, e.g., the JPEG bit sequence encoding of a photo taken at John's 30th birthday (an Digital Information Object) can have multiple concretisations as Digital Copies, i.e., in Mary's laptop, in her company's server, etc. Thus, FDOF-CM decouples encoding/formatting entities from their possible multiple concretisations, similarly to the distinction between Dataset and Distribution in DCAT [16]. In summary, FDIOs are generically dependent on FDMOs (the *materialization* relation in Fig. 3), which are themselves generically dependent on FAIR Digital Copies (FDCs). Our notion of Media Object is akin to the notion of *Information Structure Entity* in IAO; our notion of FDC is akin to *Information Bearing Entities* in IAO. FDCs will be fully developed in future work.

#### 4. FAIR Digital Object Framework Conceptual Model implementation in OWL

While FDOF can be applied to a variety of technology spaces, the Semantic Web is among the most significant ones, where OWL ontologies are extensively employed to provide semantic annotation of both data and metadata. The FAIR Digital Object Framework Ontology<sup>4</sup> (FDOF-OWL), is an OWL implementation of FDOF-CM. This implementation extends the *gUFO* ontology [17], a lightweight implementation of UFO for the Semantic Web, representing the Digital Object and FAIR Digital Object (Figs. 1 and 2) taxonomies as specialisations of *gUFO*'s classes. In this section, we demonstrate the realisability of the FDOF-CM by implementing it in OWL and using one of the real use cases.<sup>5</sup>

In FDOF-CM, one of the minimal requirements differentiating FDOs from regular digital objects is the identification by a GUPRI. However, the object identifiers in OWL are commonly used as IRIs when making statements about them. To differentiate objects from GUPRIs, allowing us to make statements about both independently, FDOF-OWL defines two distinct properties. The first, `fdof:gupri`, is a datatype property that can be assigned any data value that represents the subject's GUPRI. The second, `fdof:isIdentifiedBy`, is an object property whose range is an individualized `fdof:GUPRI` object. These two uses, exemplified in Amazon's top-selling items dataset of Listing 1 (Lines 10 and 11), allow for independent statements about objects and GUPRIs, the first supporting coinciding GUPRIs and IRIs, and the second supporting statements about the

<sup>4</sup>The FDOF implementation in OWL is available at <https://w3id.org/fdof/ontology>.

<sup>5</sup>The complete version of all use cases used in the demonstration of FDOF are available at <https://w3id.org/fdof/fois23-paper>.

GUPRI themselves. Notice that while FDOF-OWL has the expressivity to support both styles, their simultaneous use is not required, being only the datatype property option employed in our use cases.

```

1 @prefix : <https://w3id.org/fdof/fois23-paper/ex1/> .
2 @prefix fdof: <https://w3id.org/fdof/ontology#> .
3 @prefix fdoft: <https://w3id.org/fdof/fois23-paper/ex-types/> .
4 @prefix dct: <http://purl.org/dc/terms/> .
5 @prefix dcat: <http://www.w3.org/ns/dcat#> .
6 @prefix schema: <https://schema.org/> .
7
8 :amazonTop50Metadata {
9   :amazonTop50 rdf:type dcat:Dataset, fdof:FAIRDigitalInformationObject ;
10  fdof:gupri "https://w3id.org/fdof/fois23-paper/ex1/amazonTop50" ;
11  fdof:isIdentifiedBy <https://w3id.org/fdof/fois23-paper/ex1/amazonTop50_identifier> ;
12  fdof:hasInformationObjectType fdoft:Dataset ;
13  dct:license <https://creativecommons.org/publicdomain/zero/1.0/> ;
14  dct:issued "2020-10-01"^^xsd:date ;
15  fdof:isMaterializedBy :amazonTop50Csv .
16
17  :amazonTop50Metadata fdof:isMetadataOf :amazonTop50 .
18 }
19
20 {
21  :amazonTop50Metadata rdf:type fdof:FAIRMetadataRecord ;
22  fdof:gupri "https://w3id.org/fdof/fois23-paper/ex1/amazonTop50Metadata" ;
23  fdof:hasInformationObjectType fdoft:DatasetMetadataRecord ;
24  dct:license <https://creativecommons.org/publicdomain/zero/1.0/> ;
25  fdof:isMaterializedBy :amazonTop50MetadataTrig .
26
27  :amazonTop50Csv rdf:type dcat:Distribution, fdof:FAIRDigitalMediaObject ;
28  fdof:hasEncodingFormat <https://iana.org/assignments/media-types/text/csv> .
29
30  :amazonTop50MetadataTrig rdf:type dcat:Distribution, fdof:FAIRDigitalMediaObject ;
31  fdof:hasEncodingFormat <https://iana.org/assignments/media-types/application/trig> .
32 }

```

Listing 1: Amazon's Top 50 Dataset example in FDOF-OWL.

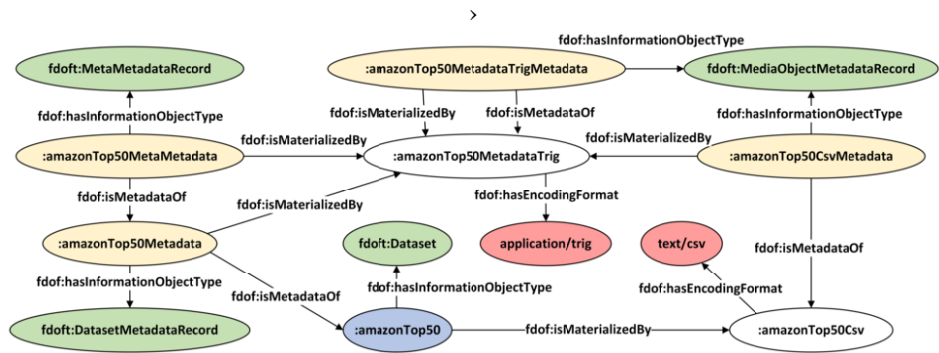
Next, FDOF-CM defines FAIR metadata records as datasets aggregating data items (Figs. 1 and 3) that record the attribution of values to objects. In the Semantic Web, data items coincide with the statements (triples) of an RDF graph. To individualise and aggregate these data items, FDOF-OWL defines no data item concept but instead makes use of RDF's named graphs [18]. Named graphs allow us to represent a set of statements (Lines 8 to 18) about a dcat:Dataset called :amazonTop50, and refer to this set as :amazonTop50Metadata. Other statements about :amazonTop50 could exist outside the named graph, but in this manner, we represent those aggregated in this metadata record, including the fdof:isMetadataOf reference back to the metadata (Line 17). Moreover, independent statements can be made about :amazonTop50Metadata, e.g., a statement about its license (Line 24) as it could also differ from the license of :amazonTop50.

As described in Fig. 2, FDOF includes a notion of FAIR information object types used to prescribe the characteristics required of a FAIR information object in a particular setting. In the context of OWL, these information object types are defined as subclasses of fdof:FAIRDigitalInformationObject, and fdof:hasInformationObjectType,

a sub-property of `rdf:type`, is used to connect the information object to the type characterising it as a FAIR (enough) FDO. In Listing 1, `:amazonTop50` and `:amazonTop50Metadata` have relations to hypothetical FAIR information object types `fdof:Dataset` (Line 12) and `fdof:DatasetMetadata` (Line 23) which could set requirements beyond having a GUPRI, such as having explicit licensing information through `dct:license` statements, and to include the publication date of a dataset through `dct:issued`.

Finally, FDOF-OWL includes the object properties `fdof:isMaterializedBy` and `fdof:hasEncodingFormat`, both related to objects of type `fdof:FAIRDigitalMediaObject`. The `fdof:isMaterializedBy` property, connects an information object to the digital media object where it is materialized, and the `fdof:hasEncodingFormat` property connects the digital media object to its characterising encoding format, much like the characterisation provided by `fdof:hasInformationObjectType` to digital information objects as both instruct consumers on how to process subject digital object. We highlight statements about the media objects from Listing 1 where `:amazonTop50` is said to be materialized in a CSV file (Line 15), while `:amazonTop50Metadata` is materialized in the very Trig file where all these statements were declared (Line 25).

In Fig. 4 we summarise in a graph notation all digital objects of the dataset example used to exemplify the FDOF-CM. The complete case includes a real dataset, information objects for its metadata and meta-metadata, and information objects for the media objects materialising them and their own metadata. To improve the graph’s readability, we also employ a colour coding of the FDOF types instantiated. FDOF-OWL has been validated through the representation of real use cases involving creative works (photos with multiple materialisations) and ontologies (including both `gUFO` and FDOF-OWL itself).



**Figure 4.** OWL dataset example in FDOF. Yellow represents FAIR metadata records; white represents FAIR media objects; blue represents FAIR information objects; red represents encoding formats (IANA media types in this case); and green represents domain information object types.

### 5. Related Work

As aforementioned, the FDOF has been inspired by the notion of Digital Objects. In [2] the authors introduce naming conventions to identify and locate digital objects, a service to locate objects from their names and elements of a digital object access protocol. Two

key entities are introduced: the obvious digital object and a repository. A digital object is defined as an instance of an abstract data type composed of two elements, data and key metadata. The key metadata includes the object's identifier and other information about the object. The paper also introduces the initially expected behaviours of the Repository Access Protocol (RAP), such as accessing and depositing digital objects.

The work on Digital Objects continued over the years, mainly under the scope of the DONA Foundation [19]. A key result of this work is the Digital Object Architecture (DOA) [20]. The DOA defines three core components, namely the identifier/resolution system, the repository system, and the registry system, as well as two protocols: the Digital Object Interface Protocol (DOIP)[21] and the Digital Object Identifier Resolution Protocol (DO-IRP)[22]. The identifier/resolution system uses the DO-IRP to create, update, delete, and resolve identifiers of digital objects. The repository system is responsible for supporting the management of digital objects, including the provision of access to the stored objects through their identifiers. The DOIP is used by the repository system and allows software applications to interact with digital objects. The registry system is a specialisation of a repository system which stores the metadata of the digital objects rather than the digital objects themselves. In recent years, part of the work on DOs has been conducted in the context of the FAIR Digital Objects Forum [23]. The FDO Forum aims at providing a community structure to foster the advance in the specification and application of FAIR Digital Objects. A number of working groups have been set to address different aspects of FDOs. The work reported in the paper has been partially presented and tested in the FAIR Digital Object Semantics Group (FDO-SIG) of the forum.

The major differences between the work on the FDOF and the DO are that the latter has been designed from the ground up to follow the aspects of the FAIR principles that can be supported by technological infrastructures and applications, and the emphasis on semantic clarity by means of the FDOF-CM.

Regarding semantic models to describe digital objects, we relate our work with other initiatives such as Dublin Core Terms (DCT) [24], the Data Catalog Vocabulary (DCAT) [16], Schema.org, and the Metadata for Ontology Description and publication (MOD) [25]. These models provide classes and relations that are used to represent entities in metadata records and to provide semantic annotations to data. The FDOF-CM does not aim at replacing them. Instead, as shown in the examples on Section 4, we reuse concepts from these models. However, in FDOF-OWL we provide greater expressivity for aggregating statements into well-defined metadata records, and we lay a foundation for the definition of more refined types of FAIR digital objects tailored for specific applications.

Finally, the work reported in [26] also carries out a UFO-based ontological analysis of the FAIR principles. We see that work as complementary to the work proposed here. For example, from a FAIR perspective, our focus in this article has been on the relation between data and metadata as well as some aspects of identification. In contrast, that study conducts the conceptual clarification of all four principles. On another perspective, the models put forth here address the characterization of the relation between content, encoding/formatting and concretisation, which falls outside the scope of that paper.

## 6. Final Remarks and Future Work

In this paper, we reported the current state of the work on the conceptual model of the FAIR Digital Object Framework. In particular, we describe how FAIR Digital Objects

are identified, how they can be classified according to a basic typing system, and how the model follows the FAIR principles also in regard to metadata.

We demonstrated the utility of the conceptual model in the context of the Semantic Web by generating an OWL ontology and using this ontology to describe the objects in representative real use cases. We argue that characterising objects by their informational value, encoding format, and related entities, provides valuable information contributing to semantic clarity. We acknowledge that this approach not only brings these benefits but also introduces more instances with their consequent properties and relations, which may increase complexity. However, it is not expected that these instances, their identifiers and some of their properties would be manually created. Supporting systems can be implemented to automate this process, generating automatically these instances, identifying and setting some of their properties and, thus, making the adoption of the FDOF easier.

In terms of future work, the FDOF-CM and FDOF-OWL, presented in Sections 3 and 4, lay the foundation for the systematic characterisation and representation of FAIR Digital Objects according to FDOF. First of all, the mechanisms of representation of information objects, their metadata, and their materialisations involve a lot of nuances that have been left out of our scope, and enhancing this coverage also includes support for non-digital objects, a subject of great relevance. Second, we focused this paper on the characterisation of FDOs based on the attributions aggregated into FAIR metadata records, leaving out the characterisation of operation types present in FDOF. FDOF's operation types are fundamental to explaining to both human and software agents how and when to process information objects, in a finer level of detail than what is possible through their materialisations and encoding types. For example, in the case of the Digital Object Architecture and DOIP support the existence of basic and extended operations. While basic operations define a set of operations that every DOIP service must support, extended operations allow different services to implement extra operations that may be offered. Third, we foresee the possibility of employing automated verification techniques, such as SHACL, for the definition of FAIR Information Object Types and the verification of their instances. This should guarantee that these instances sufficiently adhere to the FAIR principles according to the requirements defined by the application extending FDOF-OWL.

Furthermore, the work on the FDOF conceptual model occurs in parallel with the work on the rest of the framework. Reports on the whole framework, on the definitions of the protocols and expected behaviours of the participant applications and services, as well as prototypes and demonstrators, are currently underway.

Finally, as previously discussed, the results put forth here are complementary to the ones reported in [26]. As a future work, we intend to explore the connection between these two proposals.

## References

- [1] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016;3(1):160018.
- [2] Kahn R, Wilensky R. A framework for distributed digital object services. *International Journal on Digital Libraries*. 2006;6(2):115-23.
- [3] Bonino da Silva Santos LO. FAIR Digital Object Framework Documentation; 2021. <https://fairdigitalobjectframework.org>.

- [4] Verdonck M, Gailly F, de Cesare S, Poels G. Ontology-driven conceptual modeling: A systematic literature mapping and review. *Applied Ontology*. 2015;10(3-4):197-227.
- [5] Guizzardi G. *Ontological foundations for structural conceptual models*. Telematica Instituut; 2005.
- [6] Guizzardi G, Fonseca CM, Almeida JPA, Sales TP, Benevides AB, Porello D. Types and taxonomic structures in conceptual modeling: A novel ontological theory and engineering support. *Data & Knowledge Engineering*. 2021;134:101891.
- [7] Guizzardi G, Botti Benevides A, Fonseca CM, Porello D, Almeida JPA, Prince Sales T. UFO: Unified Foundational Ontology. *Applied Ontology*. 2022;17:167-210.
- [8] Albuquerque A, Guizzardi G. An ontological foundation for conceptual modeling datatypes based on semantic reference spaces. In: *Intl. Conf. on Research Challenges in Information Science*; 2013. p. 1-12.
- [9] Fonseca CM, Guizzardi G, Almeida JPA, Sales TP, Porello D. Incorporating Types of Types in Ontology-Driven Conceptual Modeling. In: *Conceptual Modeling*. Springer; 2022. p. 18-34.
- [10] Fonseca CM, Porello D, Guizzardi G, Almeida JPA, Guarino N. Relations in Ontology-Driven Conceptual Modeling. In: *International Conference on Conceptual Modeling*. Springer; 2019. p. 28-42.
- [11] Berners-Lee T, et al. Uniform Resource Identifier (URI): Generic Syntax. RFC Editor; 2005. 66.
- [12] Media Types; 2021. Available from: <https://www.iana.org/assignments/media-types>.
- [13] Guizzardi G, Wagner G. Towards a logic of the ontological dodecagon. *From Philosophy to Computational Logic: Festschrift in Honour of David Pearce's First*. 2012;60:25.
- [14] Fonseca CM, Almeida JPA, Guizzardi G, Carvalho VA. Multi-level conceptual modeling: Theory, language and application. *Data & Knowledge Engineering*. 2021;134:101894.
- [15] Smith B, Malyuta T, Rudnicki R, Mandrick W, Salmen D, Morosoff P, et al. IAO-Intel: An Ontology of Information Artifacts in the Intelligence Domain. In: *Semantic Technologies for Intelligence, Defense, and Security*; 2013. .
- [16] Browning D, Perego A, Winstanley P, Cox S, Albertoni R, Beltran AG. Data Catalog Vocabulary (DCAT) - Version 2. W3C; 2020. <https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>.
- [17] Almeida JPA, Guizzardi G, Falbo RA, Sales TP. gUFO: A lightweight implementation of the Unified Foundational Ontology (UFO); 2019. <http://purl.org/nemo/gufo>.
- [18] Carroll JJ, Bizer C, Hayes P, Stickler P. Named graphs. *Journal of Web Semantics*. 2005;3(4):247-67.
- [19] DONA Foundation;. Accessed: 2023-02-11. <https://www.dona.net>.
- [20] Digital Object Architecture;. Accessed: 2023-02-11. <https://www.dona.net/digitalobjectarchitecture>.
- [21] Kahn RE, et al. Digital Object Interface Protocol Specification, version 2.0. DONA Foundation; 2018.
- [22] Kahn RE, et al. Digital Object Identifier Resolution Protocol Specification, version 3.0. DONA Foundation; 2022.
- [23] FAIR Digital Objects Forum;. Accessed: 2023-02-11. <https://fairdo.org/>.
- [24] DCMI Usage Board. DCMI Metadata Terms; 2006.
- [25] Dutta B, Toulet A, Emonet V, Jonquet C. New Generation Metadata Vocabulary for Ontology Description and Publication. In: *Metadata and Semantic Research*. Springer; 2017. p. 173-85.
- [26] Bernasconi A, García S A, Guizzardi G, Bonino da Silva Santos LO, Storey VC. Ontological representation of FAIR principles: A blueprint for FAIRer data sources. In: *35th Intl. Conference on Advanced Information Systems Engineering (CAISE 2023)*, Zaragoza, Springer. Springer; (forthcoming) .