

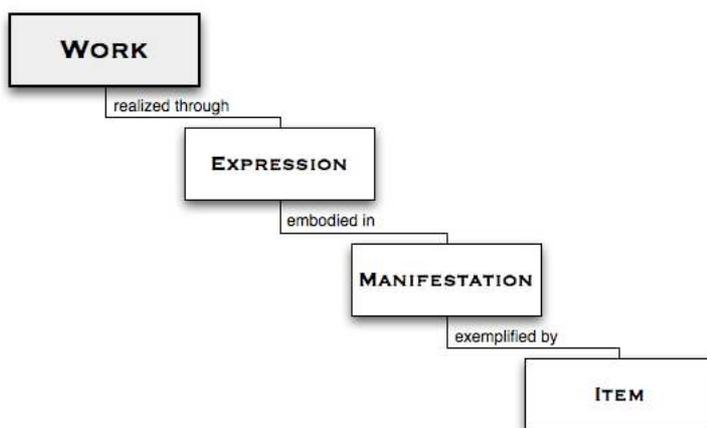
ONTOLOGIES OF INTELLECTUAL PROPERTY

Primavera De Filippi

*European University Institute, Law Department
Via Boccaccio 121, 50133, San Domenico di Fiescole (Firenze), IT*

Keywords: *Ontology, copyright, metadata*

Abstract: *As both the quantity and the variety of published material continue to grow over the years, keeping track of the different publications available and of their respective versions and editions has become a complex task. Besides, given the malleability of digital content, identifying the copy of a work consistently over time has become increasingly difficult. An ontological framework capable of identifying the different aspects of a work should therefore be developed in order to determine the extent to which the terms and conditions of a copyright license may or may not apply to the various instances of a work.*



1. FRBR

With the dramatic increase in the amount of information that has become available in the information society, a need has emerged for society to archive and organize information according to a set of established standards and principles that would allow for every piece of information to be identified and retrieved by anyone in a consistent manner. The Functional Requirement

for Bibliographic Records (FRBR) constitutes an attempt to individuate the various attributes of a work that are being commonly referred to in bibliographic records. In particular, the FRBR framework subdivides a work into four constitutive elements: (1) the work as a distinct intellectual creation, (2) the expression as a particular realization of the work, (3) the manifestation as the physical embodiment of the expression and (4) the item as a single exemplar of the manifestation.

1.1 The scope of copyright protection

Copyright law provides authors with a series of exclusive rights against the unauthorized exploitation of their works. The law does not however provide a uniform level of protection to every aspect of the work.

The work as a purely abstract and conceptual idea is not eligible for protection because the copyright does not protect the ideas underlying the work but only the expression thereof. In order to be protected, the work must therefore assume a sufficiently detailed structure which consists of more than a basic idea and which can be easily recognized in the case of non-literal copying.

Although it is generally said that copyright ultimately vests in the expression of a work, protection does not apply to the expression as the entity has been defined in the FRBR framework (according to which even a minor change in the content will inevitably produce a new expression of the work). In the framework of copyright law, the scope of protection extends far beyond one particular expression of the work. For instance, the translation of a work from one language to another or the conversion from one medium to another (e.g. the making of a movie based on a literary work) would constitute copyright infringement in so far as it entails the reproduction of the overall structure of the work.

As for the manifestation, it may sometimes be eligible for protection whenever the copyright regime provides for the protection for typographical arrangements of published editions, although this is limited to very few jurisdictions.

The correlation between copyright law and the FRBR is therefore not as straightforward as it may seem. While the expression of the work is what constitutes the ultimate unit of analysis, it is not unconditionally protected. On the one hand, copyright only protects against actual copying, so that the independent making of an identical or substantially similar expression would not necessarily amount to copyright infringement. On the other hand, copyright only protects the “original expression” of a work, so that whenever a work incorporates facts or elements that have been taken from the public domain, certain parts of the expression will not be protected under copyright law. Accordingly, in spite of its widespread application in the framework of bibliographic records, the FRBR does not appear suitable to describe the scope of protection granted to any original work of authorship.

1.2 The object of the copyright

In the context of the licensing of rights, the FRBR framework may be unable to properly determine the object of the rights that are being licensed, as well as to identify the various entities that constitute the copy of a work.

As a general rule, the item is the best candidate to qualify as the copy of a work since it is the only entity that can actually be identified in the real world.¹ In the digital world, however, the physical

¹ The FRBR framework distinguishes between the work (as a general concept), the expression (as a particular realization of the work into a specific arrangement of signs or symbols), the manifestation (as a particular concretization of the expression into a specific format) and the item (as a particular instance of the manifestation into a given medium of expression). Given that everything that subsists above the level of the “item” is fundamentally intangible, the item constitutes the only concrete entity of the whole

item has lost most of its significance to the extent that it refers to an ephemeral entity which can no longer be used to identify the copy of a work. If it cannot be identified according to its physical characteristics, a different approach has to be adopted to consistently identify a digital copy over time.

A different layer of abstraction may therefore have to be taken into account in order to identify the copy of a digital work. Yet, neither the work as a general concept, nor the expression or the manifestation of the work is likely to constitute a suitable alternative.² A solution might be that of introducing additional layers of abstraction into the FRBR framework in order to provide another set of entities that could fundamentally act as a link between the manifestation of a work and the physical representation thereof.

As such, the notion of a digital copy would no longer be regarded as a 'token' but rather as a 'type' which could be instantiated into a variety of items that do not have to be either physically or digitally identical as long as they can be logically identified as belonging to the same type (i.e. as long as they incorporate the same digital entity). Yet, even though it would theoretically be possible to introduce a series of new entities into the FRBR in order to produce a framework that is more compliant with the specificities of the digital environment, when it comes to identifying the actual copy of a digital work, the actual structure of the FRBR is however likely to be debilitating given that every entity may only subsist within one particular layer of abstraction.³ An alternative method of identification might therefore be necessary.

framework and is as such the only entity can actually be observed and experienced in the real world. See IFLA (1998) Functional Requirements for Bibliographic Records. IN SAUR, K. G. (Ed., IFLA Study Group on the Functional Requirements for Bibliographic Records.

² Identifying the copy of a work according to the general concept it embodies is likely to be too general, as it would not account for the copyright vesting in the different expressions of the work. Likewise, identifying the copy of a work according to the way it has been expressed would necessarily ignore the distinction between the different manifestations of the work which may as well be eligible for copyright protection. Finally, identifying the copy of a work according to its manifestation (e.g. PDF versus HTML) would ignore the fact that the same manifestation could theoretically be conveyed to the public in a variety of manners (e.g. with a different set of metadata, different technological measures of protection, etc) which are likely to affect the ultimate consumption thereof.

³ The structure of the FRBR is a taxonomy in which every node may have an indefinite number of children at the lower level of abstraction but only one parent at the higher level of abstraction. While it is theoretically possible to introduce new entities into the hierarchy, this entity would however have to be inserted between two other categories and cannot as such encompass different entities at different levels of abstraction. Yet, in view of the specificities of the digital medium, it is difficult to locate the entity that should be regarded as the copy of a digital work into one particular layer, given that different copies may actually have to be situated at different levels of abstraction.

2. IAO

The Information Artifact Ontology (IAO) is a new ontology of information entities, whose structure is based upon the principles of the Basic Formal Ontology (BFO),⁴ an upper-level ontology whose structure and design are grounded into a robust realist approach. The purpose of the IAO is to identify the basic entities that constitute an ‘information artifact’ (i.e. any piece of information that does not exist in nature but has been produced by a particular individual or machine) and to incorporate them into a formal taxonomy in order to define the relations that subsist amongst them.⁵ The basic configuration of the IAO is fundamentally composed of three core categories of entities: (1) the information entities themselves, (2) the processes they are involved in, and (3) the material bearers by which they are being conveyed to the public. These three categories are ultimately linked together through a series of connections that indicate the various relationships they each entertain with each other or with other kinds of entities.⁶

In particular, in so far as it relates to the realm of information artifacts, the core and the most interesting entity of the IAO is likely to be the Information Content Entity (ICE), which basically represents the entity that qualifies as the actual content of information. Yet, according to the degree of granularity that is being taken into account in the analysis of a particular information artifact, different entities may need to be identified at different levels of abstraction.

2.1. Universals and particulars

A universal is a ‘type’ which may only exist in so far as it has been instantiated into a ‘token’, which is generally referred to as a particular.⁷ Universals and particulars are therefore inherently related by a relationship of instantiation.⁸

⁴ The Basic Formal Ontology (BFO) is an upper-level ontology which basically consists of a series of sub-ontologies that can be regarded as a series of perspectives on reality. Each ontology represents some partition of reality into categories or universals. They are however only partial ontologies, i.e. they only represent that particular portion of reality which is visible through the particular perspective that each ontology focuses on. Besides, most of the entities defined in the BFO are abstract entities which may or may not accommodate concrete entities. For more details, see <http://www.ifomis.org/bfo/overview>

⁵ The Information Artifact Ontology (IAO) provides a framework for the identification and of different entities that pertain to the realm of information artifacts and for the description of their corresponding relationships with a particular portion of the world. For a more detailed overview of the Information Artifact Ontology (IAO), see <http://code.google.com/p/information-artifact-ontology/>

⁶ The scope and subject matter of the IAO can fundamentally be subdivided into four categories: (1) information content entities, (2) processes that consume or produce information content entities, (3) material bearers of information, and (4) relations in which one of the relata are information content entities. See RUTTENBERG, A. (2009) Introduction to Ontology. *ICBO: International Conference on Biomedical Ontology*. Buffalo, NY.

⁷ As a general rule, one subject can be attributed more than one predicate, which necessarily leads to the conclusion that both ‘a is X’ and ‘b is X’ can be regarded as true. Accordingly, although a and b constitute two distinct and separate particulars whose diversity has to be acknowledged, they nonetheless are related to the extent that they share an identical property X which constitutes a universal type. See RUSSELL, B. (1912) On the Relations of Universals and Particulars. *Proceedings of the Aristotelian Society*. Blackwell Publishing.

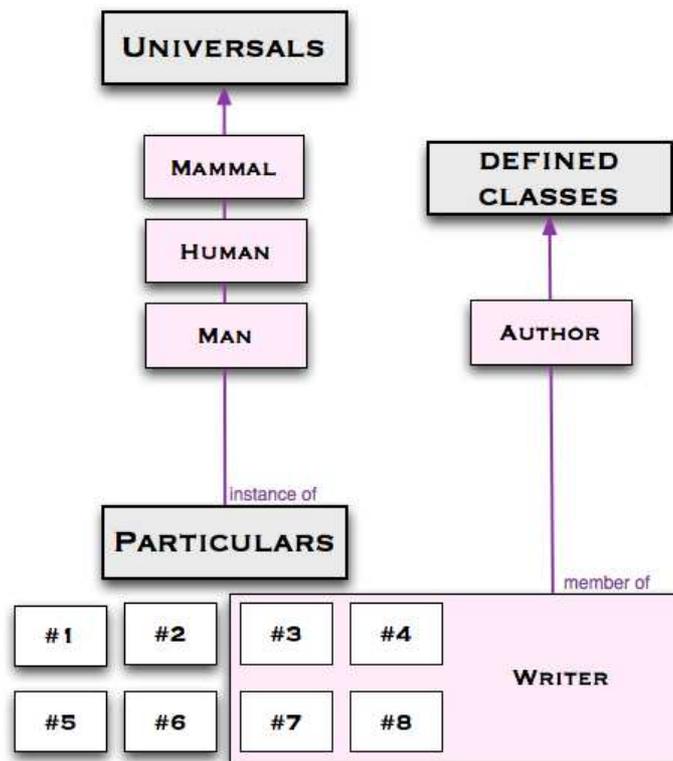
The main role of universals is to provide a basis for the classification of particulars.⁹ However, it may sometimes be necessary to introduce an additional order of classification that does not exist in nature so as to allow for different entities to be identified according to their belonging to an arbitrary categorization. The notion of a class may therefore be introduced in order regroup a series of particulars together under a common entity whose identity has been established by a series of arbitrary criteria¹⁰ that designate the specific characteristics that must be shared by every particular it comprises.¹¹

⁸ The relationship that subsists between universals and particulars is that of instantiation. For instance, the predicate 'x instance_of X' asserts a relation between a certain instance and a certain type or universal. Conversely, a relationship of characterization can generally be inferred from that particular relationship to the extent that it associates the inhering entities with their bearers. In order to exist, universals must have at least one entity that instantiates them. Accordingly, to be a particular is therefore equivalent to being one of the entity on which the universal is non-rigidly dependent. On the other hand, the relation of characterization is rigidly dependent upon the universal it refers to. Accordingly, the particular bearer of a universal quality could not exist without that given universal. See e.g. JANSEN, L. (2007) Dispositions, Laws, and Categories: A Critical Study of E. J. Lowe's The Four-Category Ontology. *Metaphysica*, 8.

⁹ The particulars that can be observed in the real world are the bearers of certain attributes or qualities that may potentially constitute a basis for classification. Universals allow for a classification of particulars that is objective, in the sense that it is theoretically independent of people judgments. Accordingly, a dog is a dog because it instantiates the universal type 'dog' and not because it is regarded as a dog by someone. Similarly, its color is black not because it appears to be black to someone, but because it instantiates the universal quality of 'being black'. See ZALABARDO, J. L. (2001) Towards a Nominalist Empiricism. *Proceedings of the Aristotelian Society*. Blackwell Publishing.

¹⁰ Classes can be defined by using a particular set of operations such as, e.g. union, intersection, complement, etc which are not available to define the boundaries of universals. Defined classes can thus be composed of a completely arbitrary set of entities which are not necessarily related to each other if not for the mere fact that they belong to the same class. For instance, the members of the class unionOf(chocolate, dog, sun) are not related to each other according to any universal concepts, but only according to an arbitrary classification. As opposed to universals, classes are generally defined not only on the basis of material properties and attributes, but also on the basis of conceptual properties that can only be observed in a particular cultural system. See e.g. READ, D. W. (1989) Intuitive typology and automatic classification: Divergence or full circle? *Journal of Anthropological Archaeology*, 8, 158-188.

¹¹ A defined class represents a particular collection of particulars which are regrouped together according to specific criteria which have been arbitrarily established (e.g. 'employee of a particular bank', 'fans of a particular band', etc) and could not be understood without taking into consideration the social and cultural context that pertains to the particular portion of reality that is being analyzed. For a more detailed overview of the differences between universals, classes and defined classes, see SMITH, B., KUSNIERCZYK, W., SCHÖBER, D. & CEUSTERS, W. (2006) Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. *Proceedings of KR-MED*.



2.2. Continuants and occurrents

Another important distinction established in the BFO is between the category of continuants and the category of occurrents. Continuants are entities that retain their identity and their form over time (such as, for instance, persons, books, ideas, etc) whereas occurrents are entities that do not possess a persistent identity or form and may only occur during a particular period of time (such as, e.g. the life of a person, the writing of a book, the formulation of an idea, etc). Common instances of occurrents are generally described as processes, events, or any other occurrence that is intrinsically bound in time.¹²

Within the BFO framework, the category of continuants has been subdivided into a series of sub-categories. On the one hand, independent continuants¹³ are concrete and material entities whose identity can be defined independently from that of any other entity (e.g. a person or a book). As

¹² According to Zemach, things are entities which are bound in space and continuous in time, whereas events or processes are entities which are bound in time but not in space given that they have no spatial boundaries. For more details, see ZEMACH, E. (1970) Four Ontologies. *Journal of Philosophy*, 67. In particular, the latter category of entities can be distinguished in two kinds: (1) entities that unfold themselves through a particular period of time, and (2) entities that constitute the temporal boundaries of the former kind of entities. See GRENON, P. & SMITH, B. (2007) Persistence and Ontological Pluralism. IN KANZIAN, C. (Ed.) *Persistence*. New York, Springer.

¹³ The IAO defines an independent continuant as a continuant that is a bearer of quality and realizable entity entities, in which other entities inhere and which itself cannot inhere in anything. Examples include an organism, a heart, a leg, a person, a symphony orchestra, a chair, the bottom right portion of a human torso, the lawn and atmosphere in front of our building. For more details, see <http://www.ifomis.org/bfo/1.1/snap#IndependentContinuant>

such, they are likely to exhibit a series of distinctive characteristics which may eventually change over time without however affecting the overall identity thereof.¹⁴ On the other hand, dependent continuants¹⁵ are entities that require the existence of another continuant in order to come into being or in order to subsist over time (e.g. an idea can only subsist insofar as it has been incorporated into a particular support, such as the human brain). More precisely, the entities that pertain to this latter category have been further distinguished into generically dependent continuants (GDC)¹⁶ – e.g. the content of a book - which can be incorporated into any kind of support regardless of the identity thereof, and specifically dependent continuants (SDC)¹⁷ – e.g. the actual patterns of ink that represents the content of a book - which are inherently dependent upon the subsistence of that specific book as a support. The main difference between a GDC and a SDC is that the former can inhere into more than one entity at a time and could potentially continue to exist even after the entity into which it has been originally incorporated has ceased to exist. As a general rule, in fact, a GDC is generically dependent upon the subsistence of at least one SDC which represents a particular concretization thereof, and that particular SDC is in turn specifically dependent upon the subsistence of the particular entity into which it inheres.¹⁸

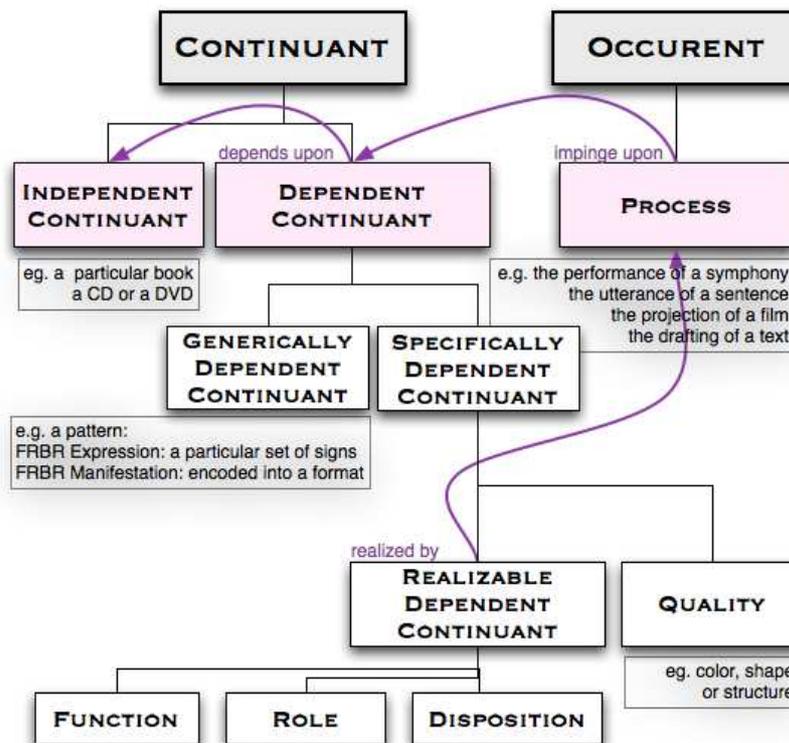
¹⁴ Independent continuants can be of very diverse nature. Yet, one of their defining characteristics is that they are such as to allow for other continuants (e.g. qualities, dispositions, functions, etc) to inhere unto them. While independent continuants can be distinguished by their distinctive attributes, the actual value of these attributes may be subject to change without actually affecting the identity of the independent continuant to which they relate. For more details, see SPEAR, A. D. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Saarbrücken, Germany, Institute for Formal Ontology and Medical Information Science.

¹⁵ The IAO defines a dependent continuant as a continuant that is either dependent on one or other independent continuant bearers or inheres in or is borne by other entities. For more details, see <http://www.ifomis.org/bfo/1.1/snap#DependentContinuant>

¹⁶ The IAO defines a generically dependent continuant as a continuant that is dependent on one or other independent continuant bearers. For every instance of A requires some instance of (an independent continuant type) B but which instance of B serves can change from time to time. An example of a generically dependent continuant is a certain PDF file that exists in different and in several hard drives. For more details, see <http://www.ifomis.org/bfo/1.1/snap#GenericallyDependentContinuant>

¹⁷ The IAO defines a specifically dependent continuant as a continuant that inheres in or is borne by other entities. Every instance of A requires some specific instance of B which must always be the same. Examples of specifically dependent continuants include the mass of a cloud, the smell of mozzarella, the liquidity of blood, the color of a tomato, etc. For more details, see <http://www.ifomis.org/bfo/1.1/snap#SpecificallyDependentContinuant>

¹⁸ A SDC can fundamentally be regarded as a bridge between a particular GDC and any given independent continuant upon which is has been incorporated. Accordingly, although a GDC can potentially be instantiated into an indefinite number of entities, it may only be materialized by the means of a particular SDC that basically represents the concretization of the GDC into a particular entity upon which it specifically depends. While the SDC is specifically dependent upon the existence of the particular entity to which it refers, the GDC is thus only generically dependent upon the existence of at least one entity into which it inheres. For more details, see SPEAR, A. D. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Saarbrücken, Germany, Institute for Formal Ontology and Medical Information Science.



2.3. Information artifacts

An information artifact should not be confused with a physical artifact. In the context of the IAO, in fact, an information artifact has been defined to include not only the information bearer as the physical medium of expression (independent continuant), but also the information carrier as the specific properties of that medium that incorporate the information (specifically dependent continuant) and the information content entity as the particular pattern that constitutes the actual body of information (generically dependent continuant).¹⁹

2.3.1. Information content entity

Defined as an entity that is generically dependent on some artifact and that stands in relation of ‘aboutness’ to some entity,²⁰ the Information Content Entity (ICE) is the pattern that constitutes the content of any given piece of information.

An important requirement is that an ICE must have been intentionally produced, in the sense that it must have been created either by a sentient with the intention to produce an information artifact, or

¹⁹ Although there is no official definition of what constitute an information artifact for the purposes of the IAO, the term has been loosely defined as “a dependent continuant or its bearer that is created as the result of one or more intentional processes.” Examples include: uniprot, the english language, the contents of this document or a printout of it, the temperature measurements from a weather balloon, etc. For more information, see the project home page at <http://code.google.com/p/information-artifact-ontology>

²⁰ An information content entity has been defined by the IAO as “an entity that is generically dependent on some artifact and stands in relation of aboutness to some entity”. See http://purl.obolibrary.org/obo/IAO_0000030

by a machine which has been intentionally designed for that purpose.²¹ Therefore anything created by a natural phenomenon may not qualify as an ICE in spite of the fact that it may actually provide information to a potential receiver.

Another distinctive characteristic is that in order to qualify as an ICE an entity has to be related to (i.e. be about) something that actually exist in the real world.²² Under the IAO framework, reality has however been given a very general definition so as to be able to include not only tangible things that can be observed in the physical world, but also universals, general beliefs, ideas, or concepts that subsist in the mind of people.²³

Finally, given that it can only be observed in the real world after it has been concretized into a particular medium of expression, an ICE does not actually stand on its own.²⁴ The subsistence of an ICE necessarily relies upon the existence of at least one independent continuant that constitutes the particular medium of expression by which the information is conveyed to the public. Yet, the extent that it can be reproduced in many perfect copies and on a variety of different media, an ICE does not actually depend upon a specific medium of expression and is therefore to be regarded as a GDC as opposed to a SDC.²⁵

²¹ According to the IAO, information entities are necessarily intentional, in the sense that they must be produced by a sentient, intentionally, or by a machine made for that purpose. This limit the scope of the IAO to exclude any kind of information not created by man (such as e.g. information created by wind, or random patterns on a beach), even when it could actually be regarded as a method of communication (such s e.g. bee waggle dances, or ant pheromone communication). From the perspective of the IAO, therefore, communication is a broader class of processes than the actual transmission of an information entity. See RUTTENBERG, A. (2009) Introduction to Ontology. *ICBO: International Conference on Biomedical Ontology*. Buffalo, NY.

²² The IAO recognizes a basic relation that relates an information artifact to a particular entity (is_about). An ICE is about another entity whenever there exists a portion of reality to which every concretization of the ICE specifically refer to, whether because it specifically or materially denotes it, or because it does at least mention one or more entities that pertain to that same portion of reality. For more details, see http://purl.obolibrary.org/obo/IAO_0000136

²³ It is however arguable to what extent is it possible to stretch reality in order to accommodate a particular entity into the realm of information content entities. Indeed, any given instance of a particular work of authorship could be regarded as being about the work as a general concept, which would basically subsist by virtue of the fact that people recognize it as a separate entity in their mind. However, it is unclear from the definition of the IAO whether it is possible for an ICE to actually stand in relation of aboutness to itself or whether it necessarily has to be about another entity.

²⁴ Information entities do not stand on their own. They do not amount to independent entities but are ultimately dependent upon the subsistence of (a) the entities to which they refer to (i.e. in order to fulfill the requirement of aboutness, an ICE is intrinsically dependent upon the existence of the entities it mentions or denotes) and (b) the medium upon which they are being conveyed to the public (i.e. in order for an ICE to exist in the real world, it is necessarily dependent upon the subsistence of at least one independent continuant). For more details, see RUTTENBERG, A. (2009) Introduction to Ontology. *ICBO: International Conference on Biomedical Ontology*. Buffalo, NY.

²⁵ The IAO defines an information content entity (ICE) as an entity that is “generically dependent on some artifact” in the sense that every instance of an ICE requires some instance of an artifact, but that the actual identity thereof is completely irrelevant and may actually change over time. The ICE can be instantiated into a variety of different artifacts and will continue to subsist as long as there exists at least one independent continuant that incorporates a particular instantiation thereof. For more details, see http://purl.obolibrary.org/obo/IAO_0000030

2.3.2. Information carrier

Before it can be incorporated into a new medium, an instance of the ICE needs however be produced. In order for an ICE to actually inhere into a particular medium of expression (e.g. a given piece of paper), it has to be first concretized into a particular entity (e.g. a given pattern of ink) which basically qualifies as an information carrier²⁶ by virtue of the fact that it actually instantiates the ICE.

As such, the instantiation of an ICE always and necessarily leads to the production of a particular carrier of information whose key function is to identify a series of attributes and physical characteristics which have to be assumed by a particular medium of expression in order for it to be capable of conveying the relevant information to the public.²⁷

These qualities are what actually constitute the information carrier. Indeed, in so far as it describes the way in which a particular ICE has to be materialized into the real world, the information carrier can fundamentally be regarded as that particular quality of an independent continuant that ultimately imparts the information content.

2.3.3. Information bearer

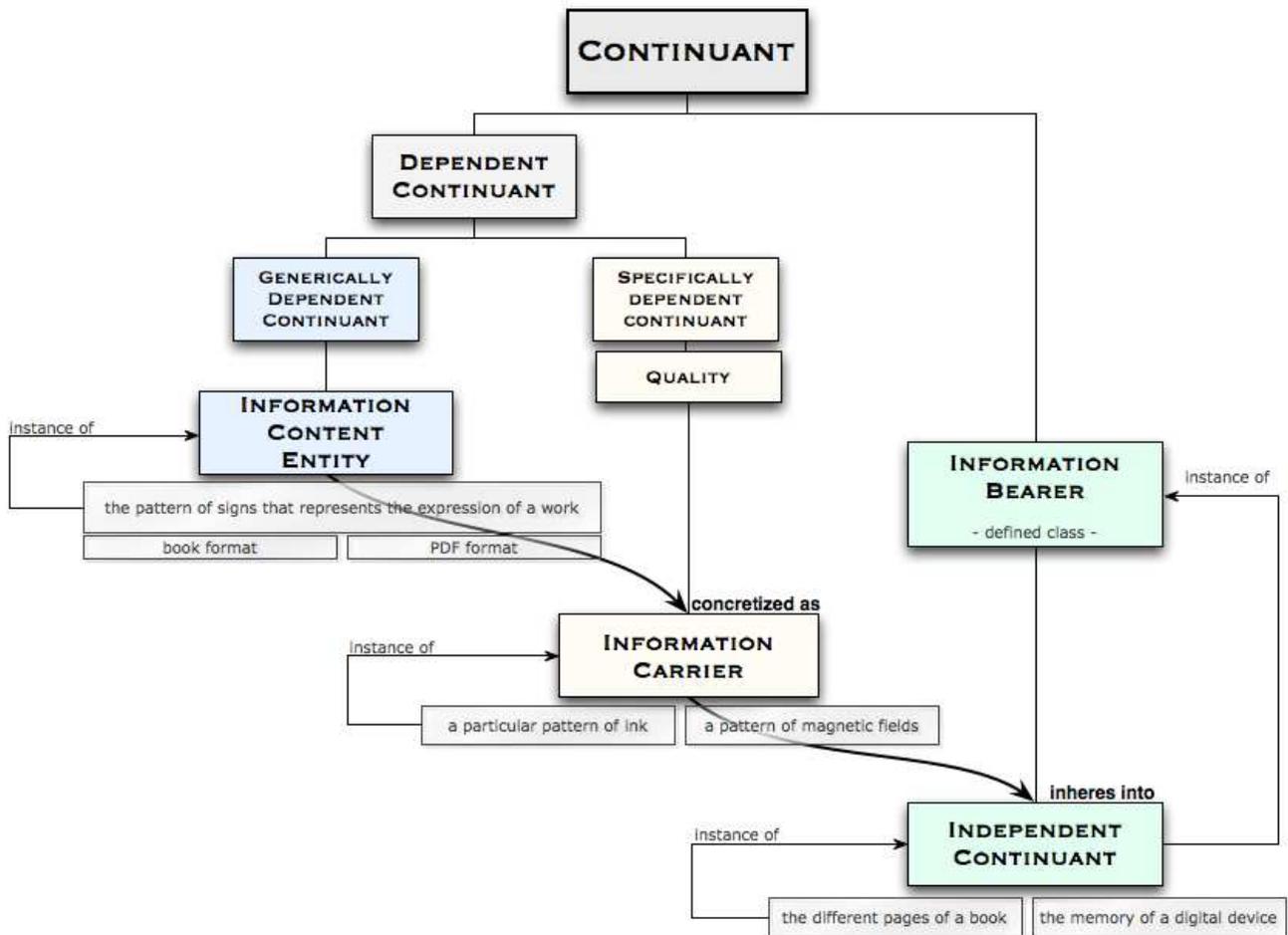
The information bearer is the physical object into which the content is encoded. It basically consists of any material entity (e.g. a book, a photograph, a CD, a hard drive, etc) upon which an ICE generically depends.²⁸

Regardless of its nature or type, any medium of expression could eventually assume the quality of an information bearer by virtue of the mere fact that it incorporates a particular type of ICE. Conversely, any medium which currently qualifies as an information bearer could eventually be deprived of this quality to the extent that it no longer exhibits the particular set of attributes identified by the information carrier (e.g. a book whose pages have been burnt, a defective CD-ROM, etc).

²⁶ An information carrier is currently defined in the IAO as the quality of an information bearer that imparts the information content, such as, for example, the pattern in ink on a page of a paperback novel. See http://purl.obolibrary.org/obo/IAO_0000125

²⁷ For instance, any piece of paper that incorporates some piece of information can basically be regarded as having been marked with a particular pattern of ink which is specifically intended to denote that information. Similarly, as soon as a digital file is saved into the physical memory of a computer, the magnetic pattern of the hard disk has to be modified in order to assume a property that basically denotes the information contained into the digital file. Besides, whenever an individual is consuming a particular piece of information, an actual mark is created into the memory section of the brain which specifically denotes the content of that information.

²⁸ An information bearer is currently defined in the IAO as a material entity, such as a hard drive, upon which an information content entity generically depends. See http://purl.obolibrary.org/obo/IAO_0000120 In other words, whenever an ICE (i.e. a particular pattern of ink) is incorporated into a physical medium of expression (i.e. a particular piece of paper) according to the specific configuration described by the information carrier which instantiates the ICE (i.e. as a result of the incorporation of that particular pattern of ink into that particular piece of paper), the physical medium will ultimately qualify as an information artifact.



2.4. The concept of a copy

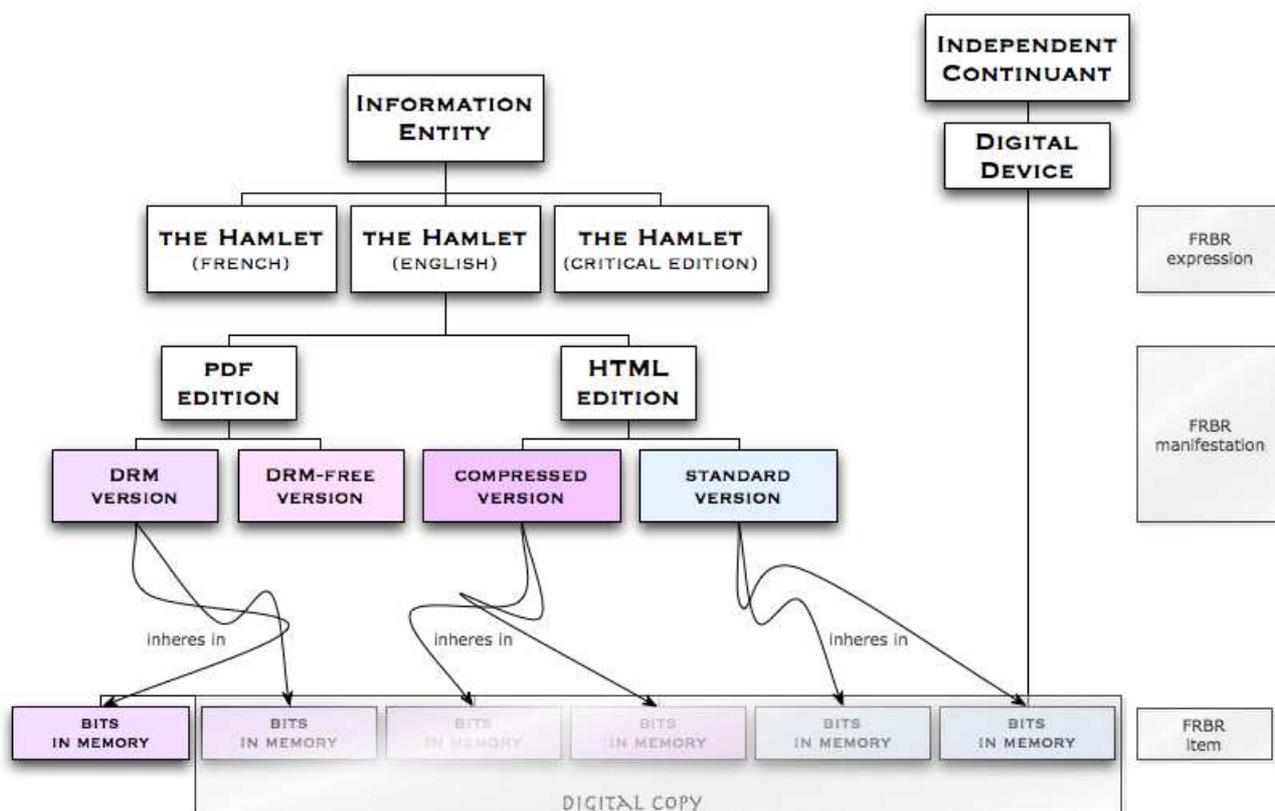
For the same reasons why the item has become unable to identify the various copies of a work that subsist in the digital environment, the notion of an information bearer as it has been described by the IAO has equally lost its value as an indicator of what constitutes the copy of a digital work.

Yet, the greater flexibility of the IAO allows for its basic structure to be continuously enhanced with a series of arbitrarily defined classes. The copy of a digital work could therefore be described by means of a defined class which would basically regroup a number of information bearers together under a single entity provided that they fulfill a particular set of criteria.²⁹

²⁹ Within the IAO framework, a defined class is an entity that fundamentally acts as a container in the sense that it basically gathers various entities together according to a series of criteria which have been arbitrarily defined. Every entity that fulfills all of the established criteria can be regarded as a member of the class, regardless of its nature or kind. A defined class is therefore not limited to one particular level of abstraction, given that it can encompass a number of entities from a variety of different levels. In the case in question, every digital copy of a work would therefore constitute a new class which would refer to all the different information bearers which actually incorporates that particular copy of the work at a particular moment in time. As such, the population of the entity is likely to change over time, as the copy is transferred from one place to another, as it is converted from format to the other, or as it is being legitimately reproduced. For more details, see *supra* Chapter 2. Section 3: The IAO approach. Subsection 4.3: The concept of a copy. Subsection B.1: Identity criteria

Accordingly, while there can be only one instance of a copy in the physical world, in the digital environment, the copy of a work may potentially be instantiated into a variety of information bearers which may or may not subsist at the same time. As such, a digital copy does not therefore qualify as an actual entity with a predetermined scope, but should rather be regarded as a generic place holder for information bearers which may eventually gain or lose their pertinence to the class according to their specific attributes of content and form.

As such, however, the digital copy does not subsist at any specific layer of abstraction. Even though the work, the expression and the manifestation of a work are actually related to each other by a directional link (i.e. the manifestation is a particular representation of the expression, which in turn qualifies as a particular realization of the work), the digital copy does not perfectly fit into that structure. On the one hand, it may in fact assume different manifestations while nonetheless retaining its identity as a particular copy of the work, whereas, on the other hand, various digital files which are identical on a bit-wise level of abstraction do not necessarily qualify as the same copy of the work.



If the function of a 'copy' is to determine the scope of the rights granted to any given licensee, it is ultimately for the right holder to determine the extent to which the various provisions of a copyright license should be applicable to the different instances of the work. Given that the 'copy' does no longer have any physical *corpus* in the digital world, only the copyright license can provide a definition of what constitutes the copy of a digital work.³⁰

³⁰ Given that the rights granted under an end-user licensing agreement exclusively refer to one particular copy of the work, the only way to establish the identity of a copy in the digital environment is to analyze the terms and conditions of the license under which it

While, in the physical world, the copy of a work is what determines the scope of application of an end-user licensing agreement, in the digital world, instead, it is the actual end-user licensing agreement that determines the scope of the copy that is being licensed, the identity of which ultimately depends upon the terms and conditions that have been incorporated into the copy itself.³¹

2.4.1. Identity criteria

Given that a digital copy may inhere into more than one information bearer, it is not possible to rely upon the physical attributes of the information bearer in order to identify the copy of a work. A different set of criteria must therefore be employed to determine equality amongst the different entities that qualify as the same copy of the work.

One condition is that all information bearers incorporate the same expression of the work. While they may not incorporate the same ICE (i.e. the same manifestation), in order to qualify as a particular copy of the work, they must necessarily refer to an identical expression of the work.

This does not mean, however, that the content may assume any kind of structure or form. Whenever a copy is produced, in fact, the copyright owner may introduce a series of requirements that any ICE has to comply with in order for the bearer thereof to be regarded as an instance of the digital copy. An information bearer incorporating a different manifestation will therefore be regarded as an equivalent copy of the work only to the extent that it complies with these predefined attributes in terms of quality and format.³²

Finally, the identity of a digital copy depends upon the way in which the public can dispose of the work, i.e. according to the terms and conditions of the copyright license (legal metadata).³³

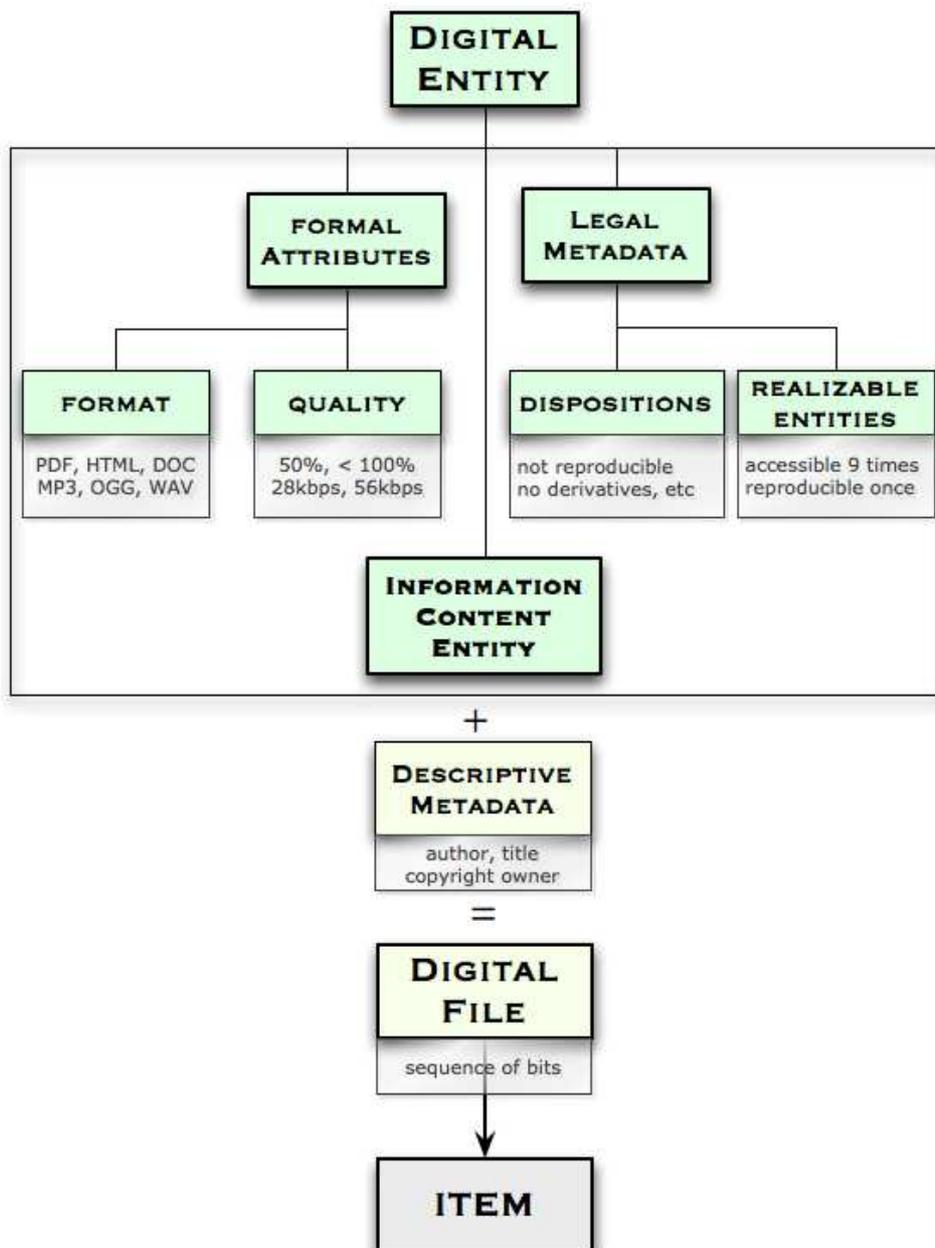
has been released. While the physical representation of a copy may change as it is moved from one device to another, and the digital representation may vary as a result of compression, encryption or conversion, the digital copy will nonetheless retain its identity over time to the extent that there exists some kind of genetic link between the various instances of the digital copy and only provided that these instances are actually legitimate (i.e. they have been produced in compliance with the terms and conditions of the copyright license). For a more detailed overview of how the concept of a digital copy could be modeled into the framework of the IAO, see *supra* Chapter 2. Section 3: The IAO approach. Subsection 4.3: The concept of a copy.

³¹ Since there is objective definition of what constitutes a copy in the digital environment, it could be said, basically, that the concept of a digital copy is ultimately determined by the definition given to the copy in the copyright license. Yet, the identity of any digital copy is fundamentally determined by the various criteria which have been established by the copyright owner at the moment in which the first instance was made. By varying the level of flexibility and the severity of these criteria, the copyright owner can therefore decide the extent to which various digital resources with different properties and attributes could actually be regarded as belonging to the same class and as such qualify as the same copy of the work.

³² For instance, the owner of the copyright in a literary work may decide to produce a particular copy of the work encoded into the PDF format and require that it stays in that format. Any other encoding of the very same work (e.g. HTML, XML, PostScript, and so forth) would therefore qualify as a different ICE which does not belong to the same digital entity. Likewise, the owner of the copyright in a musical work encoded into the MP3 format may either expect it to remain in that format, or alternatively, allow for it to be converted into a variety of different formats (e.g. WAV, WMA, OGG, etc). In addition, the right holder may however prescribe that, even though the ICE can be freely converted without affecting the identity of the digital entity, the quality of the encoding has to remain within a particular range (e.g. the identity criteria may require that an MP3 encoded with a quality of 56kbps may only be converted into an MP3 of a lower quality, or into a WAV with a quality of no more than 28kbps).

³³ Legal metadata basically consists of any metadata that refer to the legal status of a work. Various rights expression languages are being used in order to encode legal information into the particular copy of a work, in terms of copyright status, access and use

Combined together, these three elements constitute the criteria for establishing the identity of a digital copy. To the extent it fulfills these criteria, the copy of a digital work may therefore inhere into an indefinite number of digital files which may considerably differ in terms of their digital and/or physical attributes but would nonetheless refer to the same copy of the work as long as they retain the distinctive properties thereof.



conditions, permissions, rights and obligations, or any other contractual terms that may affect the legitimate exploitation of that copy. Legal metadata may or may not be used in conjunction with DRM systems in order to make sure that the various terms and conditions are actually enforced. For a general overview of how different legal data can be encoded into a particular digital work, see e.g. COYLE, K. (2004) Rights Expression Languages. Library of Congress.

2.4.2. Genetic link

Although necessary in order to determine equality between different instances of a work, identity criteria are not sufficient as such. Unless a link can be established between the various instances of a work, they cannot qualify as the same copy of the work even if they are identical in their digital representation.³⁴

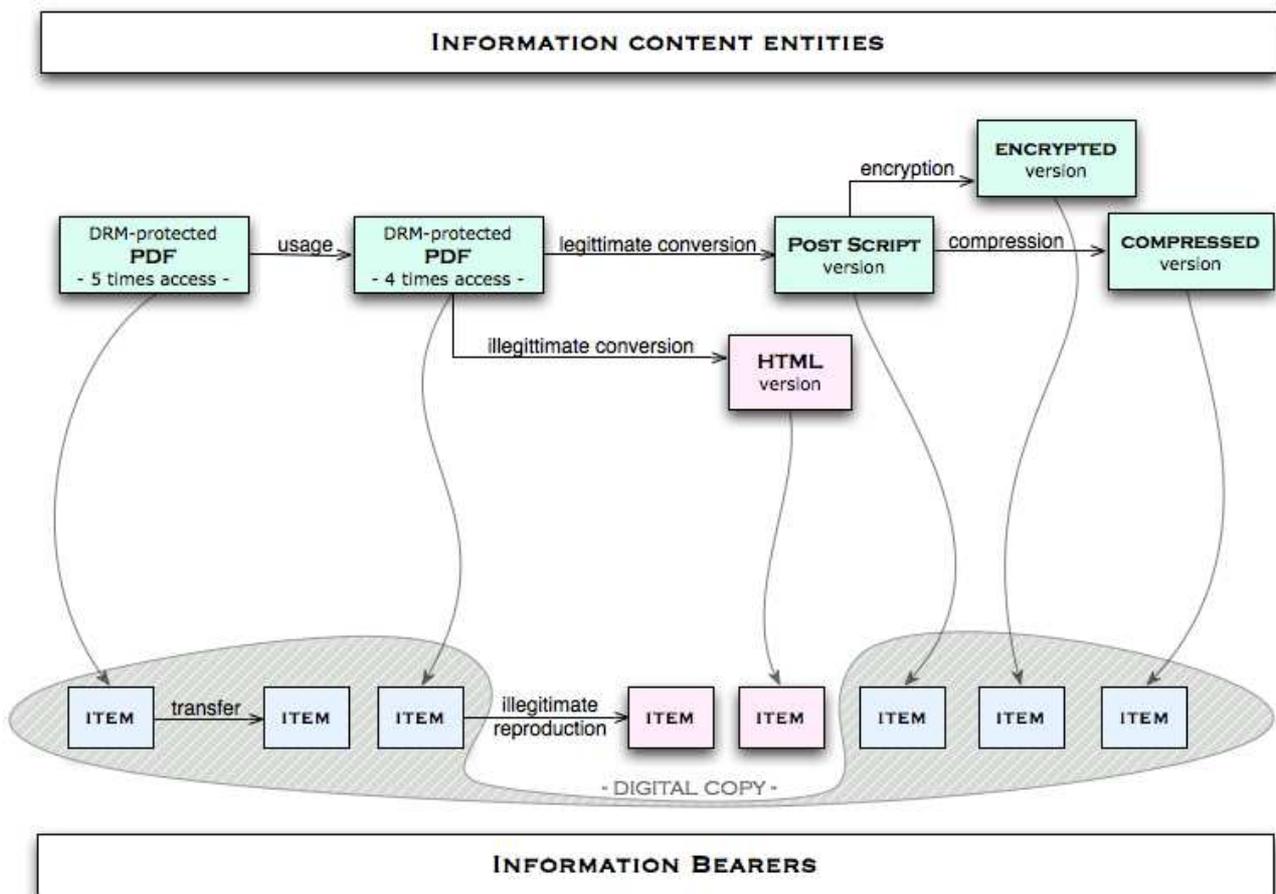
A genetic link must illustrate the various steps that separate each instance from the others. For example, whenever a copy is transferred from one device to another, the instance at the original location is destroyed and a new instance is produced which is intrinsically related to the former by virtue of the fact that it constitutes a reincarnation thereof. Similarly, the compression, encryption or conversion of a digital work from one format to another necessarily results in a new instance of the work which is no longer bitwise identical to the former but is nonetheless related through a relationship of substitution. Hence, even though the particular sequence of bits that a digital file is composed of may eventually mutate over time, by virtue of its common origin, the resulting digital file could nonetheless be regarded as an instance of the same copy.

As such, the identification of a genetic relationship is a necessary but nonetheless insufficient condition to determine the scope of a digital copy. In spite of their genetic connection, the various instances of a work may in fact only be regarded as the same copy to the extent that they each satisfy the conditions established by the copyright owner whenever that particular copy of the work has been produced.³⁵ For instance, if the license under which a copy has been released does not actually preclude reproduction, the duplication thereof will qualify as a new instance of that copy as opposed to a new copy of the work. Similarly, to the extent that it does not preclude the conversion into different formats, encoding a digital file into a new format would not result into a new copy of the work but merely into a new instance of the same copy.

In order to actually contribute to the scope of a digital copy, the genetic relationship that subsists between the different instances of a work must therefore be justified on the basis of a legitimate process.

³⁴ A digital entity essentially has the function of regrouping together a series of items which can be logically regarded as the same copy of a work. As such, the identity of a digital entity is not concerned with the physical representation of a digital work (i.e. the way in which it is represented in the physical memory of a digital device), but only with the digital representation thereof (i.e. the way in which it has been encoded into a digital medium). Yet, the digital entity fundamentally differs from the concept of a digital file in at least two ways. On the one hand, the digital entity can be much more flexible than a digital file because it is not limited to one particular sequence of bits in so far as it may assume more than one digital representation. On the other hand, however, the identity of a digital entity can be much narrower than that of a digital file to the extent that different items that incorporate the same sequence of bits may actually not all qualify as the same digital entity depending upon whether or not they all originate from the same source.

³⁵ The modification of a work at the level of the expression, the alteration or the removal of legal metadata, or the conversion of the work into a format which has not been specifically endorsed, will necessarily lead to the production of a new instance of the work which does no longer qualify as the same copy.



3. Conclusion

While both the IAO and the FRBR frameworks are capable of identifying the physical copy of a work, the structure of the FRBR is however likely to be too rigid to accommodate the various components of a copy in the digital environment.

Conversely, the IAO implements a more practical framework for the analysis of digital works and is therefore likely to be more suitable to the digital environment. In fact, in so far as the terminology of the FRBR can be imported within the framework of the IAO, the latter is likely to provide a more comprehensive mechanism to determine the object of the rights which are being licensed. Besides, through the notion of a defined class, the IAO is capable of providing an exact description of what constitutes the copy of a digital work, even though, in order to actually determine the scope of that copy, the terms and conditions under which the work has been released must necessarily be accounted for.

Accordingly, although it may never replace the FRBR in the description and the retrieval of bibliographic records, the IAO may well constitute a viable alternative for the licensing of digital works, in particular, in the context of DRM systems.

4. References

- ARANGUREN, M. E. (2005) *Ontology Design Patterns for the Formalisation of Biological Ontologies*. University of Manchester.
- BOHM, K. & RAKOW, T. C. (1994) Metadata for multimedia documents. *ACM Sigmod Record*, 23, 21-26.
- CEUSTERS, W., CAPOLUPO, M. & DEVLIES, J. (2009) *ReMINE - RAPS Application ontology*.
- COYLE, K. (2004) *Rights Expression Languages*. Library of Congress.
- DRUMMOND, N. & HORRIDGE, M. (2006) *A Practical Introduction to Ontologies & OWL*. University of Manchester.
- GRENON, P. & SMITH, B. (2007) Persistence and Ontological Pluralism. IN KANZIAN, C. (Ed.) *Persistence*. New York, Springer.
- IFLA (1998) *Functional Requirements for Bibliographic Records*. IN SAUR, K. G. (Ed., IFLA Study Group on the Functional Requirements for Bibliographic Records.
- INGARDEN, R. (1964) *Time and Modes of Being*, Springfield, Illinois, Charles C. Thomas.
- JANSEN, L. (2007) Dispositions, Laws, and Categories: A Critical Study of E. J. Lowe's The Four-Category Ontology. *Metaphysica*, 8.
- LANDES, W. M. & POSNER, R. A. (1989) An Economic Analysis of Copyright Law. *The Journal of Legal Studies*, 18.
- NEUHAUS, F., GRENON, P. & SMITH, B. (2004) A Formal Theory of Substances, Qualities, and Universals. IN VARZI, A. & VIEU, L. (Eds.) *International Conference on Formal Ontology and Information Systems*. Turin, Italy.
- READ, D. W. (1989) Intuitive typology and automatic classification: Divergence or full circle? *Journal of Anthropological Archaeology*, 8, 158-188.
- RUSSELL, B. (1912) *On the Relations of Universals and Particulars*. Proceedings of the Aristotelian Society. Blackwell Publishing.
- RUTTENBERG, A. (2009) *Introduction to Ontology*. ICBO: International Conference on Biomedical Ontology. Buffalo, NY.
- SMITH, B., KUSNIERCZYK, W., SCHOBBER, D. & CEUSTERS, W. (2006) Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. Proceedings of KR-MED.
- SPEAR, A. D. *Ontology for the Twenty First Century: An Introduction with Recommendations*. Saarbrücken, Germany, Institute for Formal Ontology and Medical Information Science.
- YEE, M. M. (1995) What is a Work? Cataloging Theorists and a Definition. *Cataloging & Classification Quarterly*, 20, 3-23.
- ZALABARDO, J. L. (2001) *Towards a Nominalist Empiricism*. Proceedings of the Aristotelian Society. Blackwell Publishing.
- ZEMACH, E. (1970) Four Ontologies. *Journal of Philosophy*, 67.