Abstract: This article presents a computer science design for the registry of documentary material based on models FRBR-ER, CIDOC CRM and FRBR-OO that is being developed in CAICYT-CONICET. The proposal leaves from the first group of entities defined by the model FRBR-ER\textsuperscript{ER} (Work, Expression, Manifestation and Item) and incorporates some of the proposals of the other two models. In general, it rescues of these models the use of object paradigm, which leads to a more rigorous definition of concepts. It takes from them, among other things, the modelization of events, which allows the representation of the documents in its temporary process. It tries to deepen in other aspects that have not been treated by FRBR-OO yet: the lack of exhaustive classification, ontological status of the Item and its relation with the physical support, the problem of the responsibility and the function of the responsible and the problem of the names. Since the presented product is being used in CAICYT-CONICET’s databases of argentine ISSN center, this article details the approach of the model to the problem of the series. Finally, some particularities of the implementation are mentioned: use of the Smalltalk language, his dialect of opened source code, Squeak, and independent frameworks that has been developed: Atón and Smallfaces.

Presentation

Opus program precedent ia a research work carried out during 2003 along with some fellows of CAICYT-CONICET: Tatiana Carsen, Hugo Garcia and Cecilia Mabragaña. Because of the classic political ups and downs of the estate administration in peripheral countries, the project was stoped just over a year. Finally, from the management of Mario Albornoz, the project has been able to leave their stage of research and now is a software product in use. Besides the author of this article, the development team is integrated by Juan Matias Burell and Hernan Morales. We want to thank the fact that they are exposing our work here to all those appointed, to Elsa Barber and the other officers and colleagues from the Biblioteca Nacional.

Before starting it is necessary to make some formal clarification. Since this model began to be thought of in 2003 following the methodology of object design -one year earlier than CRM model and four than the first version of FRBR object oriented-, some of the issues detailed in this article were developed in parallel and arrived at equivalent conclusions. Sometimes the model suffered influences that forced us to take other routes, while in other cases we have opted for different solutions and is in this latter aspect where we are going to cover in mostly. It is also necessary to clarify that most of the examples have been simplified for exhibition as it is not possible to show in detail the model within the limits of this paper.

We must also make a little clarification about the terminology. The technical names of our model follow the standard conventions for naming classes and variable in Smalltalk: the names of class,
in capital letters without space; names or messages variables, in lowercase. The graphics consists of simplified UML class diagrams: rectangles identify classes, which may contain names of attributes shown as internal rectangles, empty arrows represent inheritance relations, while full arrows represent collaborative relationships.

### Assumptions

Before presenting the theoretical and practical aspects of our proposal it seems essential to us to establish what had been the premises of which we leave. Each one deserves deepening by its own but it’s impossible to tackle at this time because of length issues. However, we will give at least some minimal justification.

The science of documentation has taken a quantum leap from FRBR model, from the mere continuation of techniques with a weak theoretical framework to the first approach of the establishment of a conceptual model from which to build a consistent theory and practice. Initiatives such as CRM and its consequence in the world of librarianship, FRBR-OO, show a tendency to find common concepts in areas of knowledge that share objects of study rather than techniques: the library and the museum. (Le Boeuf 2003a). That is why our starting point has been the definitions of the first group of entities of the FRBR-ER proposal. Much literature has shown that these concepts already existed in Luvetzky's work (Yee 2000) among others in the field of librarianship. It has also insisted that these concepts are present in an implied form not only among specialists but also among users, as evidenced by Le Boeuf in his discussion about the multiple meanings of the word "book" (Le Boeuf 2003b).

But they can also be tracked in other areas. Within the framework of structuralism and cybernetics, Abraham Moles schematize the publishing process as a channel for dissemination of thought in that way: "The creation of ideas by the author (...) [Work], the writing itself [Expression], the normalization [by typists] that destroys the uniqueness of the manuscript, the acceptance of a dissemination system [as] the book, the making of copies through the impression [Manifestation], the assimilation by individual consumers [Item] (...) "(Moles 1971).

Bearing in mind that our task is a confluence of computer science and documentation science, and that from the latter we chose the object paradigm, we must justify minimally this election, knowing that this is also a contentious issue and deserves an special discussion. Beyond the technical and economic considerations (wether their use leads to more stable products, more readable designs, sustainable, scalable, etc.) and the practical ones (wether it produces more communicable and expressive designs) we believe that the problem is a representation problem and that object paradigm allows better representation of reality, as we, humans, understand it. "A computer system should provide models that are compatible with those in the mind." (Ingalls 1981). "Traditional software structuring techniques concentrate first in function –the function of a programme- ... But human cognition often works the other way, recognizing things first, ant the functions that connect them afterward" (Liu 1996). "On one level, objet programming is more natural because it allows us to organize information in ways that are familiar to us, as illustrated in the clas hierarchies. On a
deeper level, it is more natural in that if reflects nature's own techniques for managing complexity." (Taylor 1997)

Traditional programming, which come from mathematics, tends to be more reductionist when the represented domains becomes more complex. Objects paradigm, however, uses biological and linguistics metaphors, which decreases the reductionism insofar as we move closer to the way we conceive of the reality.

The semiotic-linguistic aspect is present in this first definition of Smalltalk father, Alan Kay, as defined the Smalltalk programming language as a “medium” (XEROX 1976). "Programming involves a process-oriented computers that control entails rigid, strict, precise and linear, while 'communication' is a process aimed at involving human understanding and consensus, and often is inaccurate." The two key elements of object orientation, object and message, make evident in a clearly way this issue: "All actions undertaken in Smalltalk are sending messages to objects, and everything returns a message object" (Mortensen 2001).

Regarding the influence of biology, Kay tells us: "Philosophycally, Smalltalk’s objects have much in common with the monads of Leibnitz and the notions of 20th century physics and biology." (Kay 1993 ). The biological metaphor, are always presents in these technical keywords: inheritance, classification and behavior.

We must therefore give some definitions for object key concepts. Object paradigm has three key features: objects, classes and messages. An object is composed of some computer operations and some information, which modifies the structural design, according to which programs are built with data structures and algorithms. Objects communicate with each other by sending messages, and the response to a message is another object. The inheritance is the mechanism that allows a class to share structure and behaviour defined by one or more other classes. Classes are "factories" of objects (Liu 1996).

Since CRM and FRBR-OO models use them, we can not fail to mention a contentious issue within object paradigm realm: the multiple inheritance. Multiple inheritance cases used in these models are quite complex to use as an example and the desirability of its applicability would require at least one specific paper, so we will illustrate the problem with simple examples. If we want to represent the idea "bats, pigeons and airplanes fly, while chickens and trucks do not" in a pre-existent hierarchy such that "pigeons and chickens are birds, bats are mammals and trucks and airplanes are vehicles" using multiple inheritance we could establish an abstract entity "flying objects", from which some of these objects would inherit their properties while the others wouldn't. Instead, using simple inheritance, the objects which have the property "to fly" simply implement it without changing the original hierarchy. Some know how to respond to the message "fly" and others don't. In object terminology this is known as polymorphism: the ability of objects from different classes to have the same behavior expressed in different ways.

Our position is not rigid about it, but we tend to think that at the time it is necessary that a class inherits characteristics of more than one class, inheritance as a conceptual tool is no longer appropriate and introduces unnecessary complexity: we are dealing with the limits of classification and its proper to apply here the old teaching of William of Ockham: "entities must not be multiplied
Before addressing the treatment of our model, it is necessary to make a brief reference on the ways of representation of bibliographic information with computers. We will analyse the two traditional models and the difference that object orientation makes.

The so-called SIR (IRS) (information retrieval system), as noted Moya (Moya Anegón 1995), suffer from a lack of theoretical basis in comparison with other models of information and its justification has always been of a practical nature. However, for the purposes of this study we emphasize that the main characteristic of these systems from the standpoint of representation is that the entity or concept that we want to represent and their correlation in computer science are equivalent. The model defines a high level entity (the document) to which others are subordinate (the author, title). These entities do not have an informatic correlation; because of the limitations of the file systems used, they are represented as repeated strings. This concept could be summarized with the equation \( \text{a document} = \text{a record} \). If we want to represent the relationship "a document is produced by several people" we must repeat the string that identifies the author in each document. The problem that brings with it this repetition is, on one hand, the lack of standardization, since there will be as many strings representing an author as documents he has produced, but on the other hand, the entity "author" did not have a unique physical equivalence. The authority records come to meet this challenge: defining a table of authors and then establishing the link with the documents. But if we want to represent the inverse relationship "a person can produce several documents," we must repeat the name of the document in each author record. Instead of names, we may use more precise identifiers, but the assignment of these, considering the file system usually used, must be done manually and therefore is subject to human error.

The next step is to let the system establish these identifiers; then we arrive to another model: the relational. Normalization is the strength of this model -the most extensive in programmer's world- precisely because it poses as the first rule not to repeat any information (first normal form). Taken to its ultimate consequences, the relational model allows a full normalization: each entity will be represented by a table. To depict the double relationship "document-author" is necessary to create a third table "documents-authors": the one that contains the relationship. But now there is no distinction between entities at different levels: physically all entities correspond with the same type of representation. In other words, from the standpoint of computer files, there is no physical difference between the table for documents, the table for authors and table for author-documents relationship. The higher level entity is now the result of the dynamic operation on the tables, so the equation is now: \( \text{a document} = \text{a query} \). Knowledge of the entity is represented in one place and data in another.

Objects paradigm, while eliminating the traditional separation between data and algorithms, allows us to preserve the best of both worlds. On the one hand there are types of objects of different levels of abstraction, each equivalent to one entity to the represented domain. But as objects establish collaborative relationships among other objects, and these relations are not somewhere outside the objects, but within themselves, then the thoroughness of the relational model remains:
each document 'knows' its authors, each author 'knows' its documents. The former equation can be expressed now as: **a document = an object.**

**The OPUS model**

The first articles on the model FRBR (Noerr 1998, Velucci 1997) proposed a model based on the inheritance among entities of Group 1 which has been criticized by several authors (Carsen 2003, Renear 2006). In these initial ideas, the use of the term "inheritance" is not taken from the object paradigm and its meaning is not always clear. It should be noted that in recent FRBR-ER the idea of inheritance is much stricter and delimited.

In our model, the four entities of Group 1 of FRBR-ER correspond to four classes called **DocumentalWork, DocumentalExpression, DocumentalManifestation** and **DocumentalItem**. The class **DocumentalWork** knows all its expressions, while each **DocumentalExpression** knows its work. This double reference is repeated at the other levels, so no matter what the access point, all levels can be traversed. Clearly the relationship between the entities is a composition one rather than inheritance one.

In the FRBR-ER final report, although it no longer speaks of "inheritance", the main entities suffer from a lack of unbundling, as has been observed several times, in some cases to talk about lack of "types" ("Perhaps FRBR lacks a “type” attribute for each of the three upper entities" [Le Boeuf 2003b, p. 11]), other times mentioning specifically the class inheritance ("The overall class of publications [Manifestation en terminologia FRBR]. NM] can be divided into several media (books, films, etc)." [Heaney 1995, p. 141]).

This problem derives partly from the fact that it's an entity-relation model, which must be finally transferred to a relational database, resulting in unmanageable complexity, and also to the dependence on traditional cataloging practices, which point to the description of the entities through the assignment of strings rather than the definition of entities in themselves and their
relations with the context. Almost all the attributes of the entities identified in Group 1 FRBR-ER
denounce this problem. Some examples: the attribute Form of a Work which is defined by
extension giving examples such as novel, play, poem, essay, symphony, etc., (IFLA 1997) have
no meaning in an object design; it is impossible to establish a proper behavior on -say- a
symphony, because we would be talking on a much more comprehensive entity, the work. The
question "Which instruments are performed in the play?" could be addressed to a novel. The
attribute of the Original Language of Work could lead us to ask "In which language is the work?" to
a sonata.
In the model FRBR-OO this problem is not treated, as this work, which is still in a state of
development, has been stoped before. No subclass has been defined on the entities of Group 1 for
representing "types".
In our model, DocumentalWork class is an abstract class that shares its behavior with any work
and it provides specific subclasses for each type in particular, a process that finally derives in the
so-called documentary types. Each of the subclasses has its own attributes and behavior. For
example, the attribute language only makes sense in subclasses of DocumentalWork in which its
expressions contains text (TextualWork), while the attribute key, designed exclusively for the
repertoire of classic-romantic western musical tradition, will make sense in instances of
MusicalWork subclass, and so on.
This subclasificación process can continue on other levels, to the extent that makes sense in each
case. For example, the attribute medium of performance in expression, will have meaning only in a
DocumentalExpression subclass that conceptualize temporary arts.
Other usual distinctions could be designed following this process in the other entities of Group 1.
Take another example: the traditional distinctions between editing and printing or reprinting are
defined on the one hand with the subclass of DocumentalExpression, BibliographicEdition, and
another with the subclass of DocumentalManifestation, BibliographicPrint.
It is important to remark that in all these cases, when we talk about attributes we are not talking
about strings, or pointers to tables, but about objects; this means that when we speak, for example,
about the attribute language, we mean that the object itself -say-, an edition -knows an object of
*Language* class that owns the behavior associated with the concept of language. For the sake of readability these relationships are not represented in the pictures.

The reverse problem has not been taken into account: the lack of abstract superclasses covering shared behaviors. In the process of object design this need not always arises in the first moments, when the goal is to represent the specific entities that are evident at a first glance. However, once the behaviors of the most visible entities are established, repetitions of shared attributes and behaviors emerge, which indicate the need for a more general abstraction that simplifies the design and eliminate overlaps.

While CRM model is prodigal in abstract superclasses, we cannot say the same about FRBR-oo. At least in the first three levels of group 1 entities we can see certain elements present in all them:

1) Name: the uniform title of the work, the title of the expression (ie, the title that the author wanted to put that is not necessarily the final title), the proper title in the manifestation.

2) Identifiers: possible universal identifiers for works and expressions that most likely will be established in the near future, ISSN, ISBN, etc., in the manifestations.

3) Whole-part relationship: a multi-part work, multi-expression for its part, clusters of works in a manifestation, such as anthologies, and so on.

4) Relation between other entities: the relationships between the entities of the group 1 that have been detailed in the IFLA FRBR final report (IFLA 1997) and other works (Velucci 1997).

This deserves the existence of an abstract superclass *DocumentalEntity* representing the document before belonging to a certain level. Please note in the graphic that objects of class *DocumentalEntity* establish whole-part relations and other relationships with other objects of the same class, and therefore any of its subclasses. This means that relations between the same levels (work-work, expression-expression, etc.), as the combinations of relations between different levels (work-expression, manifestation-expression, etc.) are supported. In the full model these relationships are clearly delineated by the same design processes shown here, but we cannot dwell so much.
The problem of the item and physical carrier

Another problem that we considered insufficiently treated by FRBR is the ontological status of the item and the relationship with physical support. Both FRBR-OO and FRBR-ER consider the physical object as part of the documentary world. For example, the report FRBR-OO in the definition of class F10 Object said: "This class comprises items of material nature that are units for documentation," while the definition of class F5 Item is: "This class comprises physical objects (...) that carry a F41 Publication Expression (...)" (IWG 2006, CIDOC 2004).

However we see the item as a new entity arising from the action of the cataloguer on a physical object when she incorporates it to a collection of documents.

Certain properties of the item, such as its itinerary along different collections -locations where it was, dates of admission and discharge, institutions or people who owned- are not properties of physical objects, but only considering them as documents. While this itinerary impacts the physical object in the form of stamps, annotations, etc., each of these marks talks about the status of the object as a collected, sorted and inventoried document, rather than about its physical history.

Moreover, not every physical object that is represented in a documental system has to be necessarily a document. To take one example that only a naive gaze can consider remote: in a virtual library system designed to represent a library in its physical dimension, furniture, shelves, rooms, etc., are not documents, but they should be represented also and possess the same attributes as physical objects that books or CD-ROMs.

Once it's clear that these entities are at least partly outside the strict domain of documentation, it is not advisable to use the subclasificación: entities as "book", "disk", "journal" are also part of the physical world and as physical objects share many properties with generic physical objects, but from the documentary point of view may differ greatly; changes suffering in their interaction with the physical universe are subject to the same laws. Therefore, our model defines two distinct classes, DocumentalItem and PhysicalObject related by composition rather than by inheritance.

Establishing identity between physical object and item leads to another problem: not every item matches a single physical object. For example, a music recording on CD-ROM represents, from the standpoint of a collection of documents, a single item. However, as physical object it is composed by an optical disk, a booklet and a plastic box, all of them physical objects with different properties, which may suffer several changes. Only if the first is missing or damaged, the work is no longer accessible; if the second is not present, there is a loss of referential information. If only the plastic...
box is missed, from the information point of view the damage is negligible: it can be easily substituted.

This scheme leaves out very a important distinction that is clearly established in the relationship manifestation-item: mold and copy. Physical objects can have this relationship, being the mold an "ideal" version form which arises the copies. Therefore, the right thing would be that each manifestation had a relationship with the "ideal" physical object, which we called PhysicalObject, and each item for a particular physical object, a PhysicalObjectCopy which in turn is a copy of the mold (the PhysicalObject).

The properties present in all copies (weight, size, material, etc.) appear in PhysicalObject class, while the particular changes which have suffered each copy, make up the state attribute of PhysicalObjectCopy class.

We leave for another occasion the consideration of electronic documents, which, although they share some characteristics with physical objects can not be considered as such.

**The problem of responsibility and function**

Something similar happens with the issue of responsibility. In FRBR-ER entities Person and Corporate Body are seen only in their role as participants in the documentary process, following the tradition of documentation standards. This problem has been observed repeatedly and even the FRANAR proposal note this when it clarify: "The person entity does not reflect a person exists in the real world, but an intermediate between the real world and the universe in the catalogue" (IFLA UBCIM 2007). Although the same terminology is retained, still has some ambiguity.

A user of an information system is also a person. Why should a person has attributes such as "documents"? This would force us to define a class User, which share almost all the attributes of a person, but in a totally independent hierarchy and that would be a grave error in design.

The FRBR-OO proposal, on the other hand, being a harmonization project with CRM, "inherits" many of its virtues, and in this case breaks with the documentary bias when it derives Family F6 and F7 Corporate Body from class E21 Person, representing the concepts of family, corporate entity and person, without considering their role in the documentary universe.

FRBR-OO "inherits" also the event outline of CRM to address the issue of documental responsibility at several levels, which ensures that the registration in time of the documental process is ensured, an omission of the original FRBR that was refered several times about. (Heaney 1997)

However, the problem that arises immediately when talking about documental responsibility is problem of the functions: how do we reflect the fact that a person is the author, translator, interpreter, etc., of a work? There is no mention in the two FRBR proposals to this issue.

We have taken the scheme of events mentioned, but taking care on the fact that these models charge all the behaviour associated with documental responsibility on events (F30 Work Conception, F31 Expression Creation, F33 Identifier Assignment, etc.). Instead, we preferred to emphasize the person who participates in the action rather than in the action itself. If we wanted to represent the "translator" according to FRBR-OO, we should subclassify F31 Expression Creation
and arrive to a class *Translation*. Instead we preferred that each responsibility function has a class associated with the person rather than action, in other words, we prefer a *Translator* class to a *Translation* class.

What we have modeled is a class that represents a person considered as responsible of documents, *DocumentalResponsible*, which delegates in the class *Person* the behavior associated with a person, but which also knows relationships and attributes of that person in their interaction with specific documents. Any *DocumentalEntity*, i.e., a work, an expression or a manifestation, knows its responsibles; inversely, a *DocumentalResponsible* knows their works, expressions or manifestations. Through the mechanism of delegation, all *DocumentalEntity* can communicate with the person itself, so this relationship is not lost.

This class also allows solving the problem of the different functions of the responsible. Now the function of being the author, publisher, translator, and so on, is not in the person, because one person can play several roles in different documents. Each *DocumentalResponsible* represents a particular person, who has a particular function in certain documents.

With this approach it is possible to define characteristics of a particular function without resorting to inappropriate attributes. For example, for register instruments executed by a musician in a given work, the class *MusicalInterpreter* can define a particular attribute *instruments* for that purpose.

As we mentioned below, we have adopted the model events of CRM and FRBR-OO; using the terminology of our model it could be summarized as follows: an entire series of events could happen to any *DocumentalEntity*, which consist of actions taken on the document in a particular place (*Jurisdiction*), a certain date (*Date*) and conducted by an agent (a *DocumentalResponsible*). The model provides several subclasses of events but we cannot dwell more.

**The problem of names**

Another problem related with responsibility, but that covers almost all the entities in the documental world is the problem of the names, which we mentioned in our previous work (Carsen 2003). We
have not room now for detailing how CRM and FRBR treats names and how are they treated in our proposal. However we wish to highlight an important aspect of the model to give a clear idea on how is the final schem regarding the responsibility question. The relationship established between DocumentalEntity and DocumentalResponsible is not direct but is mediated by name. This means that a document has as responsible one of the several different names that a person can use. That way both the original inscription of the document, i.e., the statement of responsibility, and the relationship between the person and the document (using the mechanism of delegation already spoken) are retained. This solves several problems associated with personal names without producing repetition: the problem of pseudonyms, variants of names, married names, etc.. The same method is used to relate the events with places: the place of publication of a document is one of the names that have (or had) a particular jurisdiction (a city in this case).

The problem of the series

A particularly critical aspect of our task has been the representation of series, since one of the collections from which we must address is that of the national ISSN center. The series in FRBR has so far had little treatment in specific literature, and in some cases there is no agreement on how should be the implementation of work level in a serial (Berg 2004). Our starting point has been the one that enunciated Pat Riva (Riva 2003) regarding the relationship between serials and FRBR entities: The work is the underlying series as a whole, the expressions include the text in the original language, the various translations, audio versions, etc., manifestations consist of the original printing, reprints, their productions on microfilm, and their versions on CD remote, while the item is the full subscription. This view seems to be shared by, among others, Yee's assertion that "Users surely consider both the database and the journal they seek (under any title it has held) as different versions of the same work." (2003 Yee ). We have set aside for the moment the controversy regarding the possibility of a superwork level that many scholars see usefull for serials, in the belief that the problems that have raised this "extra" level can be solved by establishing appropriate relations among the four basic levels, following the idea of serial family (Riva 2003).
Among the many problems raised by the modelling of serials and bibliographic materials in general, the issue of meetings has a special place. Against the tradition of considering meetings entities as documents, FRBR\textsuperscript{-oo} has distinguished clearly that the congresses are independent events that may have some connection with documents, when it includes them as examples of class \textit{F11 Event}. But the model stops exactly there: a congress is just an example of that class.

In our model we have defined a class \textit{Meeting}, whose instances may (or may not) be related with documents, both series or any other kind of document. As subclass of class \textit{FrequentEvent}, it inherits the attributes \textit{frequency} and \textit{period}, as it incorporates the attributes \textit{languages} and \textit{responsibles}. Note that the latter attribute is a collection of objects of class \textit{GenericPerson}, which affirms the desirability of a clear distinction among entities in the design; we reuse a class, something that we could not have done if persons had kept attached to their role of documental responsibles. The object person in this case refers to the person or persons responsible of the meeting, not the responsibles (authors, publishers, etc.) of the document.

**Implementation**

We would like to give some implementation details of software Opus. Because our work is framed within object paradigm, we used the programming language that represents it the best and that does not incorporate any elements of other paradigms: Smalltalk. Within dialects Smalltalk, we have chosen Squeak as it is an open source program. Our work has produced, besides the application itself, some frameworks that will soon be available to the programmers community: Aton, wich permits interact with ISIS databases and Smallfaces, an extension of Seaside framework for building Web interfaces.

To keep as much as possible within the same working method, for object persistence we have opted for an object database. Taking into account the budgetary constraints that we are undergoing, we opted for a free object database: Magma.

To this object database we are migrating the existing records from ISIS databases of CAICYT: the database of the ISSN national center, the database Union Catalogue of Serials and the databases of the libraries Docsa, Gietz and REMCyTA.

Currently this application is in use at the ISSN national center.

**Bibliography**


