UNIT 1
CONCEPTUAL MODELING (EER)

BASES DE DATOS 2023/2024
CFGS DAW

WORKSHOP A: FIRST STEPS WITH ER SOLUTIONS

Reviewed by:
Sergio Badal
Author:
Paco Aldarias
Date: 09/20/23

SUGGESTED SOLUTIONS

SOLUTION EXERCISE 1

The first thing we will do is to look for entities. We could think of "Authors" and "Books". However, how many authors are we going to store? We would only have one occurrence (also called instance): Andrea.

Is it worth saving in this case? Probably not. She is the one who is going to manage it and she already knows herself, she does not need to store information about herself. Is the entity "Books" needed? Of course it is, it is where she is going to store the information of the books she writes.

Secondly, look for the relationships. In this case we only have one entity and it doesn't seem we need reflexive relationships, so in this case we have no relationships and therefore we don't have to look for either participation or cardinality.

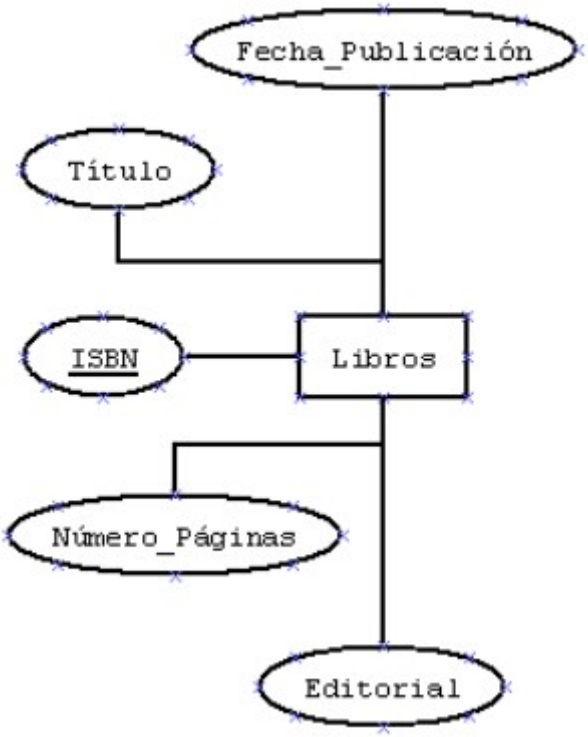
Let’s go now for the attributes (also called fields or properties) of the entity. As the statement does not specify anything special we will place the most usual ones for this entity. However, this information should be checked with Andrea (the client), in case she wants to add something else or remove any that does not interest her.

Some of these attributes could be: ISBN, Title, Number_Pages, Publication_Date, Publisher. Remember you are only entitled to set an identification/main key attribute if no more are stated.

Now that we have of all the attributes, which one do you think can uniquely identify each of the books?

For a single author (as it is the case here) we could think that the "Title" could be a good candidate because it does not seem logical to publish two books with the same title, however what happens if you review the book three years later and make a new reprint? You could not use the same title, you would have to change it, since the main key cannot be repeated.

That’s why we consider that the ISBN is the main key field, since it is a set of characters that you are assigned when you publish a book and it is unique for each book. If the book were to be republished, it would get a new ISBN when it is published, so there would be no problem.



At this point, maybe you think we could split the Publication_Date into day, month and year but **AS A COMMON RULE** we don't split the dates into fields.

SOLUTION EXERCISE 2

The context of our problem has been modified and, with it, our E-R model will also be modified. We perform the same process as in the previous exercise:

First, we look for the entities. Now we will be interested in having the entity "Author" since there are several authors that will be stored in that entity (it will have several occurrences) and on the other hand we also need the entity "Book" to store the information of the published books.

We look now for the relationships. The relationship will be between the entity "Author" and the entity "Book" and we can call it "has_written".

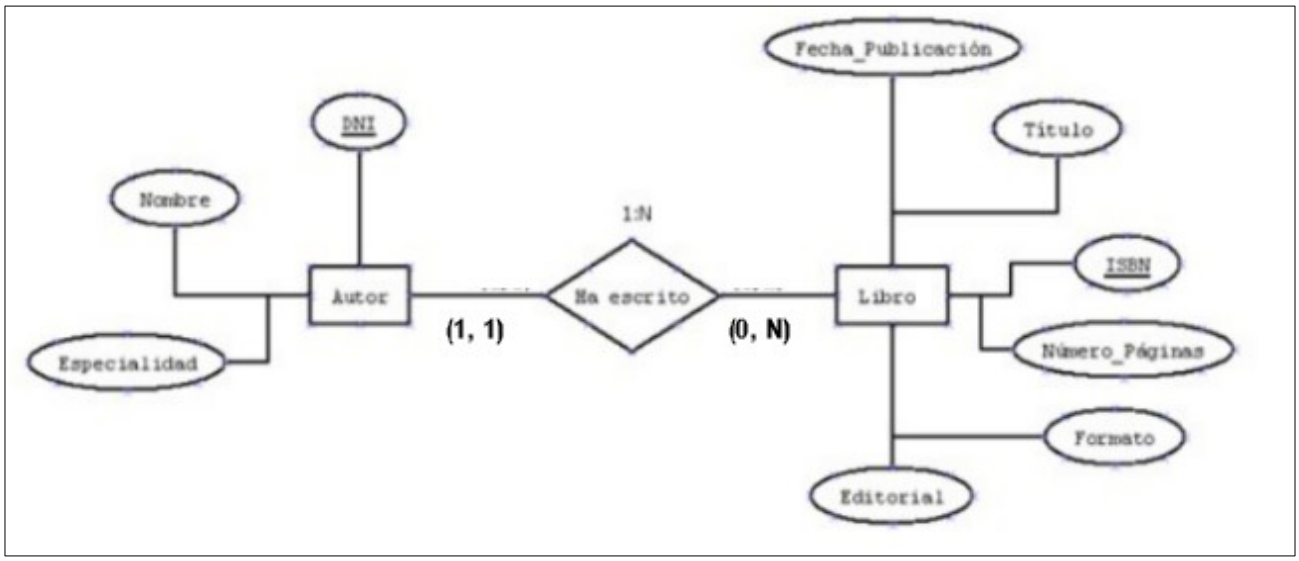
We go now for the participation of this relationship. An "Author" CAN writte several "Books", then it would be (0,n). On the other hand, a "Book" has always been written by an "Author" and at most by only one, since the statement does not say that a "Book" could have been written by several authors at the same time (in case of doubt **we would have to ask the client** or confirm it by looking at the physical occurrences of the "Books"), then it would be (1,1). Then the cardinality (the maximum of each participation) would be 1:n.

Next, we look for the attributes of each entity. Some of the attributes of the entity "Author" could be: ID, Name and Specialty. Remember you are only required to set an identification attribute if no more are stated.

As a key for the entity "Author" we will use the field DNI (id_card) which is unique for each of the occurrences/instances of the entity.

Some of the attributes of the entity "Book" can be: ISBN, Title, Page_Number, Publication_Date, Publisher, Format.

The primary key (or identification field) of the entity "Book" will be ISBN, as we show before.



AS A COMMON RULE, we will write all entities in singular "Subject", "Book"... or all in plural. Either option is irrelevant as long as the same criteria is applied to ALL entities in the same diagram.

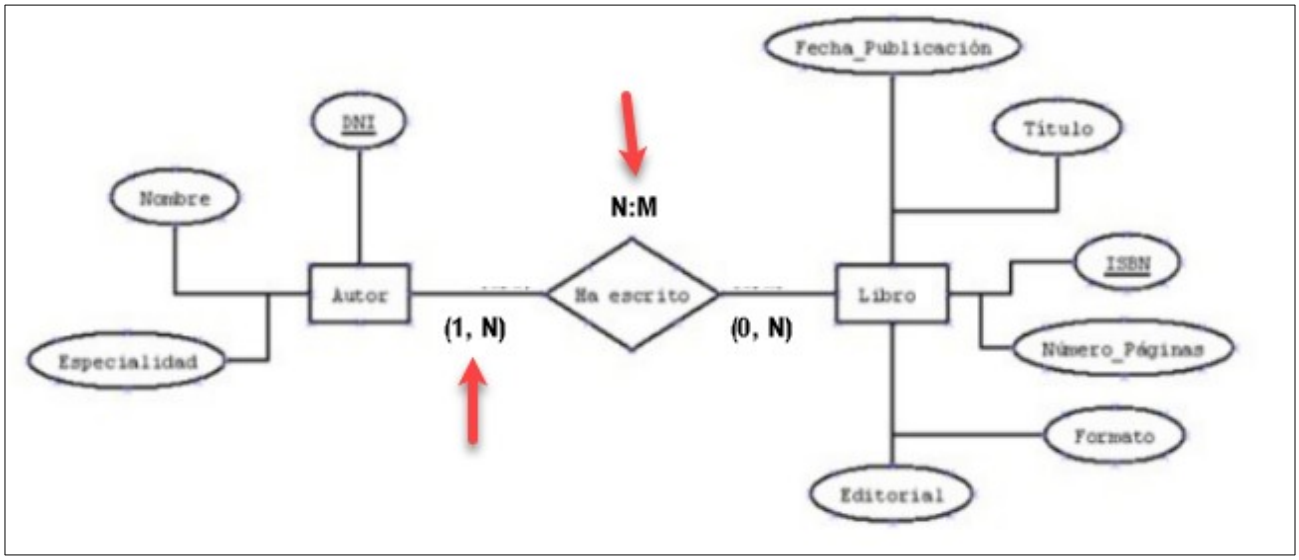
SOLUTION EXERCISE 3

Does the model we have from the previous example work for this new context?

To the first question we have to answer that the same model of the previous example does not work, because one of the conditions of that model was that each book was written by a single author, however now the possibility arises that the same book can be written by several authors, so the participation and the cardinality will change.

If anything should be changed, what would it be

To the second question (already partially answered in the previous one) we have to say that the participation of “Books” in the relationship “has_written” changes. In this case a Book will have been written by at least one author and at most by several. So, it would become (1,n). Therefore the cardinality will also change and we will have N:M.



SOLUTION EXERCISE 4

Be aware the statement talks about books, publications, works and texts. If you notice, they refer to the same entity. This will happen often and you have to be careful not to create new entities.

This new context opens a wide range of possibilities, although we are going to make the simplest one, we could include more entities and relationships that now are not going to be very productive, so they are not included and become attributes of another relationship. For example, "Center" to which the author belongs, "Course" in which a certain subject is taught, etc.

Initially, the entities we are going to work with will be the ones we already had, "Author" and "Book" and a new one, which are the subjects. For consistency, we will write it in singular: "Subject", although we could have put them all in plural.

The relations that we are going to have are the ones we already had between "Author" and "Book" (has_written) and a new one between "Book" and "Subject" ("Is suitable for") that will allow us to establish a correspondence between the books stored in our DB and the subjects for which they are suitable.

We place ourselves in "Book" and look at "Subject": We look for participation and cardinality. A "Book" "**is** suitable for" a "Subject" (1,1) and as the statement says we will find that for a "Subject" there **can** be several suitable "books", then we will have a participation (0,n). Taking the largest of each pair we will obtain the cardinality (1:n) as shown in the figure.

We look for the attributes (we could include some more if we consider it necessary).

- For the Author entity we add the attribute "Center".

- For the entity Books there are no changes.

- For the entity Subjects: Studies, Course, Name, Number_Hours_Week. We will see later what will be the primary key.

With the attribute "Studies" we refer to whether it is ESO, BACH SCIENCES, BACH SOCIALS, CFGS ASIR, CFGS DAM, CFGS DAW, CFGM SMR, etc. With "Course", the first, second... fourth (for ESO) and the "Name" is the name of the subject.

Now we have to find the main key, we could think that the most appropriate is "Name", however, it may happen that the subject of Mathematics is called the same for

any course of ESO and for the courses of BACH, then as it is not unique it cannot be the main key. What other options do we have?

The attribute "Studies" cannot be the main key, because ESO has many subjects, so it would not be unique for each occurrence. The attribute "Course" cannot be the main key either, because we will have many first year subjects and there will also be first year of ESO, of CFGS DAM, etc. The attribute "Number_Hours_Week" cannot be the main key either, since there may be several subjects that have the same number of hours per week.

Therefore, we have found that none of the attributes alone can be the main key, so we will have to try a composite main key, i.e. formed by the combination of several attributes.

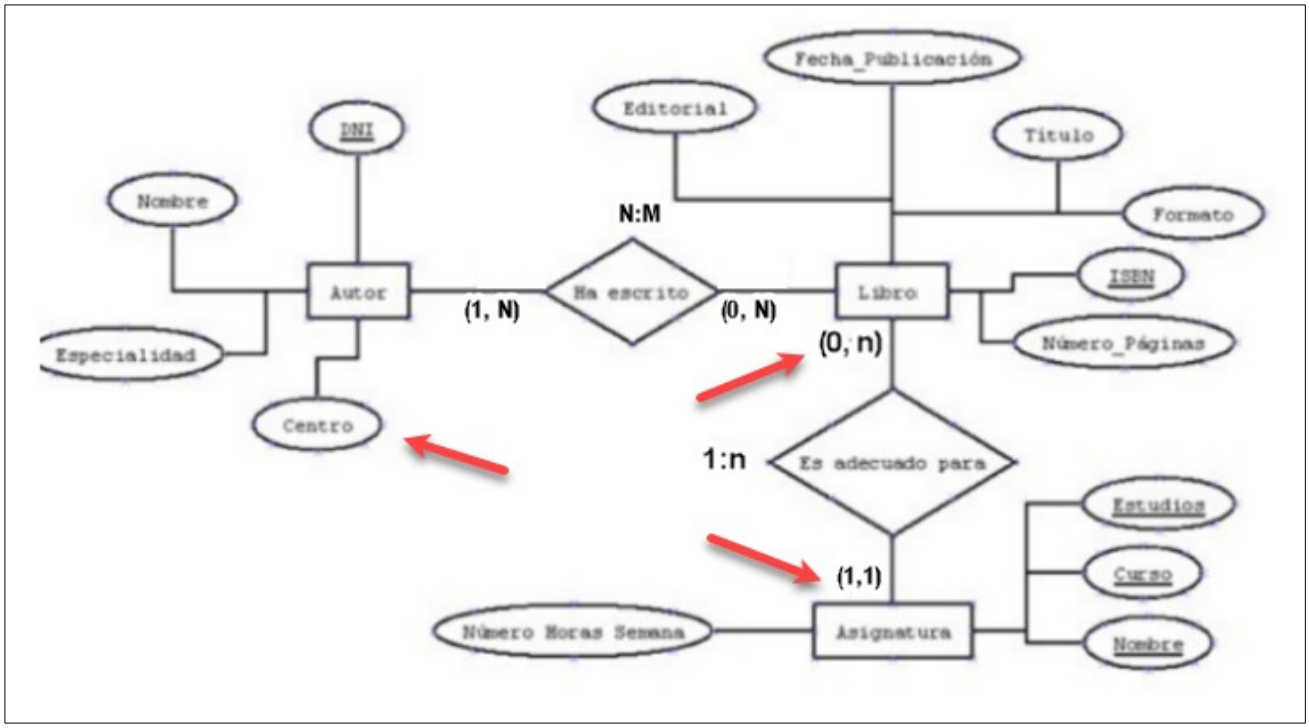
If we use "Studies" together with "Course" we could have for ESO + 1ESO, ESO + 2ESO, etc., In each of these courses there will be multiple “Subjects”, so it is not valid because it does not univocally identify each of the occurrences of the entity “Subject”.

If we use "Studies" together with "Name" we could have: ESO + Mathematics, ESO + Spanish, ESO + English, but it is normal that both Mathematics, Spanish and English are used in the four courses of ESO, so it will not uniquely identify each occurrence of the entity “Subject”.

The other alternative is to use "Course" next to "Name". In this case we would have: Primero + Mathematics, Primero + English, etc. However, what would happen with the English of the first year of ESO, with the English of the first year of “Bachillerato de Ciencias” and with the English of the first year of DAM? It also does not help us to uniquely identify each occurrence of the entity “Subject”.

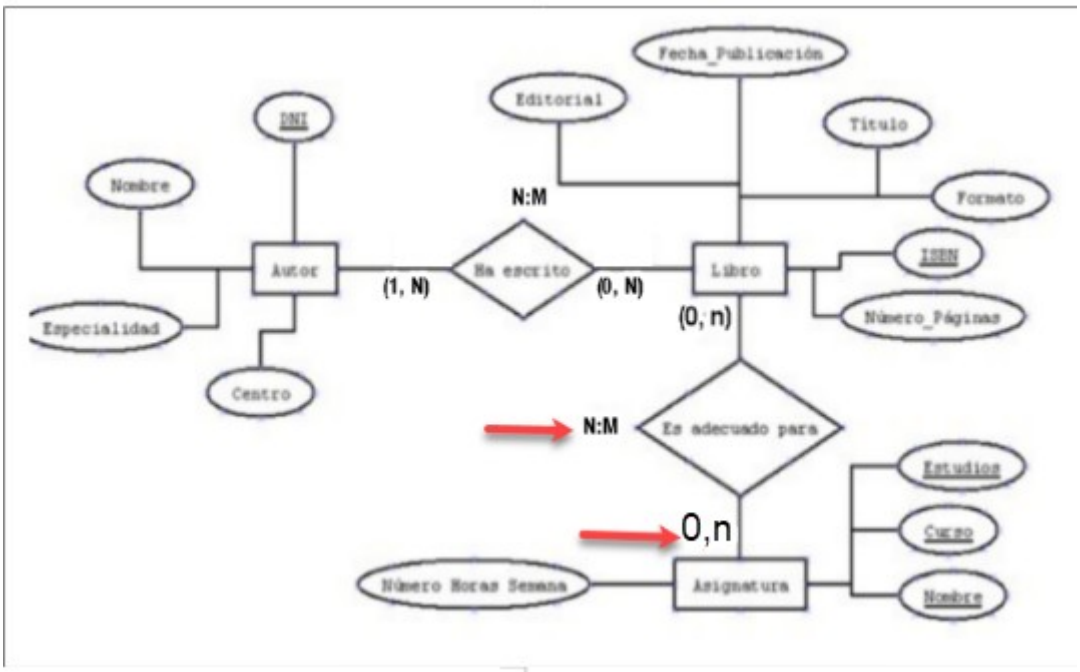
However the main key is necessary, then we have to look for it by joining more fields. In this case "Studies", "Course" and "Name". We can see that there will be only one ESO + Primero + English or one CFGS DAM + Primero + English or one “Bachillerato de Ciencias” + Primero + English.

Now, each "Subject" is uniquely identified by the union of these three attributes. Therefore, this will be our triple main key: "Studies", "Course" and "Name"



SOLUTION EXERCISE 5

With the solution of EXERCISE 4 this possibility is not contemplated since the participation between "Book", "Is suitable for" and "Subject" is (1,1). That is, a book can only be useful for one subject and in the statement we have found that it can happen that the same book can be useful for several subjects. It would therefore be a (0,n) participation and would also change the cardinality of the relation to N:M.



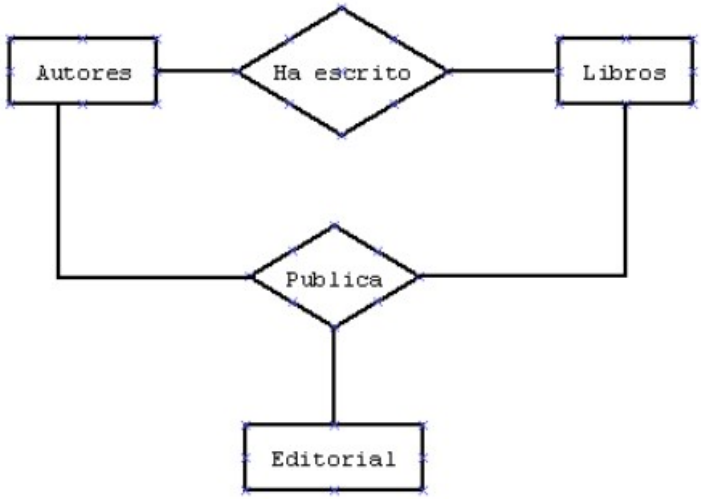
SOLUTION EXERCISE 6

Note: From now on we will also avoid (to simplify the graphics) drawing the attributes in the ER model, although we will document them when making the model. This is very common in academic environments and in the previous modeling phases in professional environments.

First, we look for the entities that will form our E-R model. In this case we find the entity "Authors" (remember that there are books that have been written by several authors), "Books" and "Publishers".

Then we look for the relationships. We can find the relationship “has written” between "Authors" and "Books". We can also find another relationship “publishes” between “Authors”, “Publishers” and “Books” that shows us the correspondence between a publisher that has a publishing contract for a book with one or more authors of that book.

Our initial diagram could be something like this:



We have a relationship of degree 2 (binary) and another of degree 3 (tertiary) that joins “Authors” and “Books” entities. **Since there’s a loop, what we have to ask ourselves is if they are redundant**, that is, if with only one relationship (the binary or the tertiary) we have all the information we need, or if each one provides different information, and in that case both will be necessary.

Let's study the case. The relationship “has written” tells us which author or authors have written a certain book. On the other hand, the “Publish” relationship tells us which

author or authors have published a book with a publisher. Obviously, the authors who publish the book with the publisher are the same ones who have written the book.

So this “Publish” relationship provides us with the same information that the relationship “has written”, then **the latter is redundant and can be removed from our model.**

Now we are going to calculate the participation in the relation of degree 3 that we have obtained (**remember you can set x as the minimum participation in tertiaries**).

• Given a specific “Author” related to a specific “Publisher”, he may have published several books with that publisher, then the **Books’s entity maximum participation will be** n => (x, n).

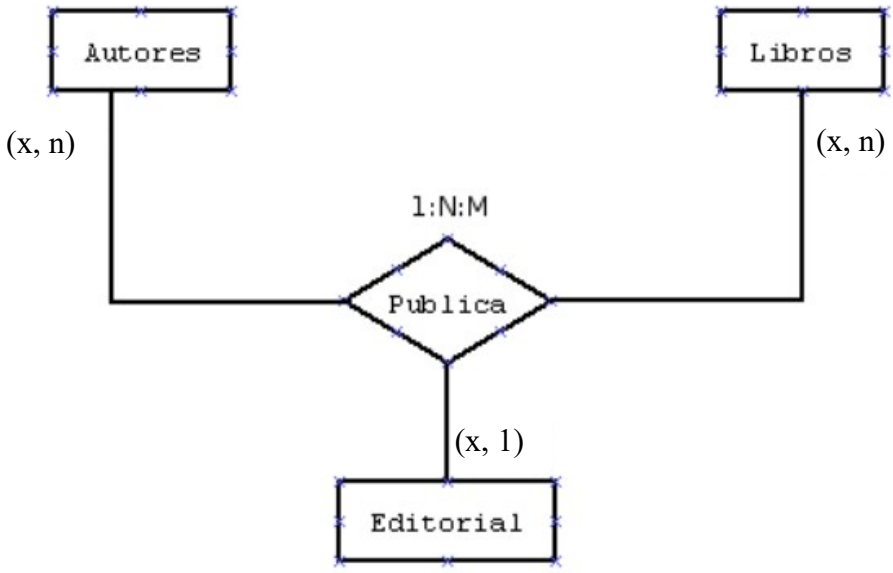
• Given a specific “Publisher” related to a specific “Book”, it may have been published by several authors with whom it has made contracts, then **Author's entity maximum participation will be** n => (x,n)

• Given a specific “Book” related to a specific “Author”, it can be related at most by one “Publisher”, as indicated in the statement, so **Publisher's entity maximum participation will be** 1 => (x, 1).

Therefore, the cardinality obtained will be 1:N:M.

Staying as follows **(being more correct with “Editorials” in the plural)**:

The next step will be to look up the attributes of our entities and relationships.



- For the “Authors” entity: DNI, Name, Specialty, Center => Primary key: DNI
- For the “Books” entity: ISBN, Title, Number_Pages, Format => Primary Key: ISBN
- For the “Editorials” entity: CIF, Name, Telephone, Mail => Primary key: CIF

One more thing. If we want to save information about the date of the publication contract, where would that attribute be placed? And if we want to store information about the commission percentage that each author will receive for the sale of a copy of a book, where will we store that information?

Well, both the date of the publication contract and the percentage of the commission of each author are issues that go into the publication contract that is drawn up for each author of each book by each publisher, if you look at the three parties involved they are associated by the Public relationship that will be the appropriate place to store these two attributes that the statement asks us for.

We’ll see attributes in relationships with more detail next week.

(being more correct with “Editorials” in the plural)