Assumptions

- [*School*, *Journal*, *Publisher*, *Series*, *Author*, *Editor*] have multiplicity 1:N in their relationship with *Divulgation*, as we believe it doesn't make sense for a database entry to not be connected to any other entity
- The correlation between *Journal* and *Publisher* is checked before database upload; they are then logically connected in the diagram through the *Publication* entity
- A *Thesis* (be it *PhD* or *Master's*) can only be submitted at one school, where it was written

==========

ER Description

The process of creating our diagram began with the identification of the main entities revolving around what we knew about a bibliography.
We, therefore, started modelling entities such as authors and editors, as well as scientific journals with their respective publishers. All of these are logically connected to a general Divulgation entity representing any kind of scientific work.
This rough sketch was then expanded and elaborated upon, with some research and looking at the DBLP database, by adding universities and series and all the divulgation types, such as thesis, books and articles.
We then finalised the diagram by adding a primary key and all the required attributes to the modelled entities.

Notes

- While the DBLP database contains a *Cite* entity with an array of keys referring to the cited records (*Divulgation* entities), we believe a more "ER-oriented" approach is to have a *Divulgation* directly cite another entity of the same type with an N:N multiplicity
- The *Divulgation* entity only has one mandatory relationship, the one with the *Author*(s); all the other relationships have a 0:N multiplicity, given the fact that they are not shared by all the different kinds of *Divulgation*
- For the *Divulgation* entity, we chose the ee (electronic edition) attribute to be its primary key: the DOI was initially chosen, but seeing as not every document has an assigned DOI, we chose to go with the more general link to the document's resource
- A similar situation to the one above happened with the Author entity: while the initially chosen primary key was the Orcid, a unique identifier for every researcher/author of scientific works, we decided to use a more generic incremental integer ID: this choice was made for the database to be compatible with decades-old publications, whose authors did not have an Orcid and might have never gotten one
- Both the "belongs_to" relationships are to be read from *Inproceedings* (*Incollection*) to *Proceeding* (*Book*); the reason for choosing the 0:N multiplicity on the side of the *Proceedings* (*Book*) instead of 1:N is that these two could exist even without being a collection of titled sections
- An incremental ID was chosen for the [*Editor*, *School*, *Publisher*, *Series*, *Journal*] entities as well since their only meaningful attributes for the sake of this database might not be unique (such as a name)
- The url attribute of the *Divulgation* entity describes the location of the document in the DBLP website, where we got the database entries from
- The crossref attribute of *Inproceedings* (*Incollection*) refers to the *Proceedings* (*Book*) entity containing it