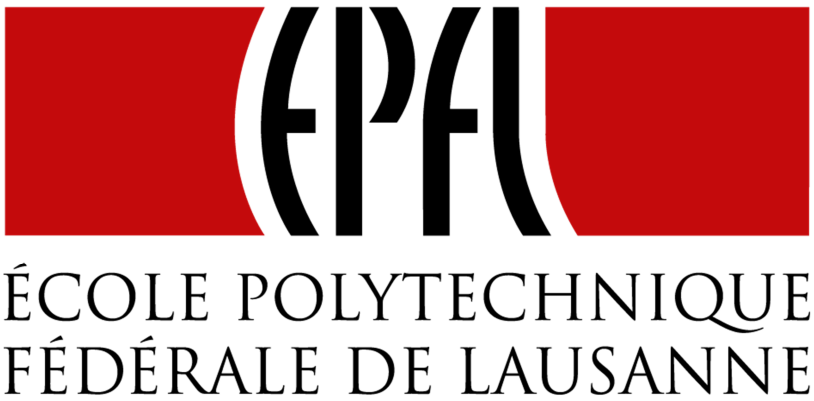
Introduction to Database Systems

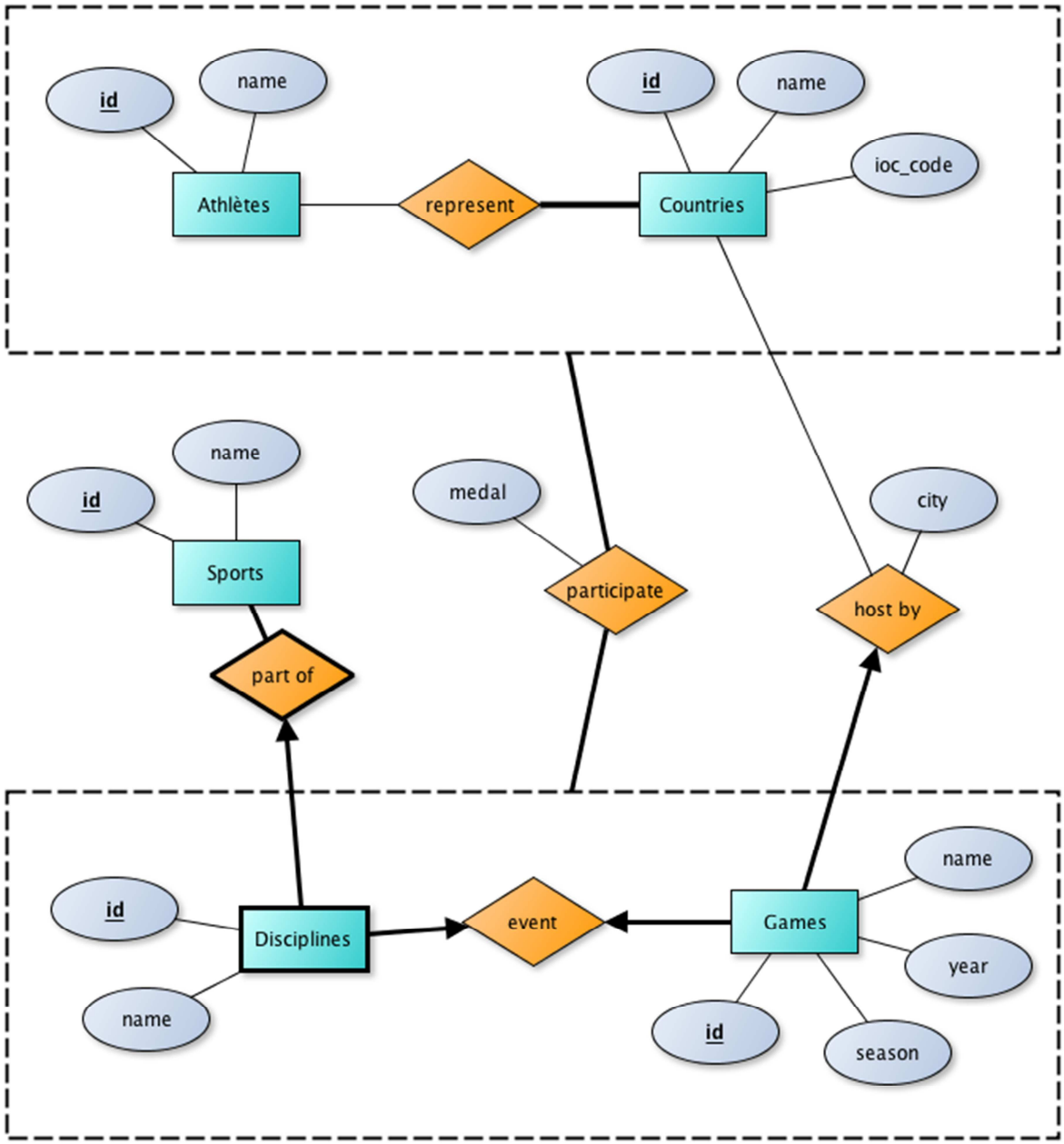
Olympic Games



Bastien Antoine (203267)
Denoréaz Thomas (183785)
Dieulivol David (185078)

Academic years 2012-2013
(March 23, 2013)

I. New ER Model:



From the analysis of the Dataset, here are our assumptions:

- An **Athlete** can represent only a **Country** for a **Game**. However, he can represent another **Country** for another **Game**.
- A **Game** can only be hosted by one and only one **Country**, but this **Country** can host several Games.
- Each **Discipline** is defined by its **Sport**.
- An *Event* is characterized by only a **Game** and only a **Discipline**.
- A *Participant* is formed by both a *Representant* and an *Event* and has an attribute about the medal.

II. Changes since deliverable 1:

After the first deliverable, we have made some simplifications to our model.
There are still two aggregations standing for a representative (**athlete** and **country**) and an event (**discipline** and **games**). These aggregations are bonded by the relation *Representant_participates_Event* which models the participation from a representative to a **discipline**.
We have removed the other relations between them because there is only redundant information and we can put the medal attribute in the participation relation.

III. Queries:

Here are some explanations of queries that seem difficult to understand:

- The query A is the intersection of athletes who won medals in summer and who won in winter.
- The query C is selecting the minimum year (so the first event) where a country won its first medal. It returns for each country the corresponding Olympics which mean the host city and the year.
- The query D is selecting the union of the best country (most of medals) of all of the winter Olympics and the best one of all of the summer Olympics.
- The query G is taking for each Olympics the maximum of all counts of participants in each country.

IV. Importing data:

The main issue that appeared while importing the data is that we cannot have the discipline for each athlete. This is problematical for the relation *Representant_participates_Event*. There is information in the csv files about the sport that practices each athlete but the sport cannot define an event. So we have decided not to set the discipline as a primary key so that all non-medalists can still be stored in the database.