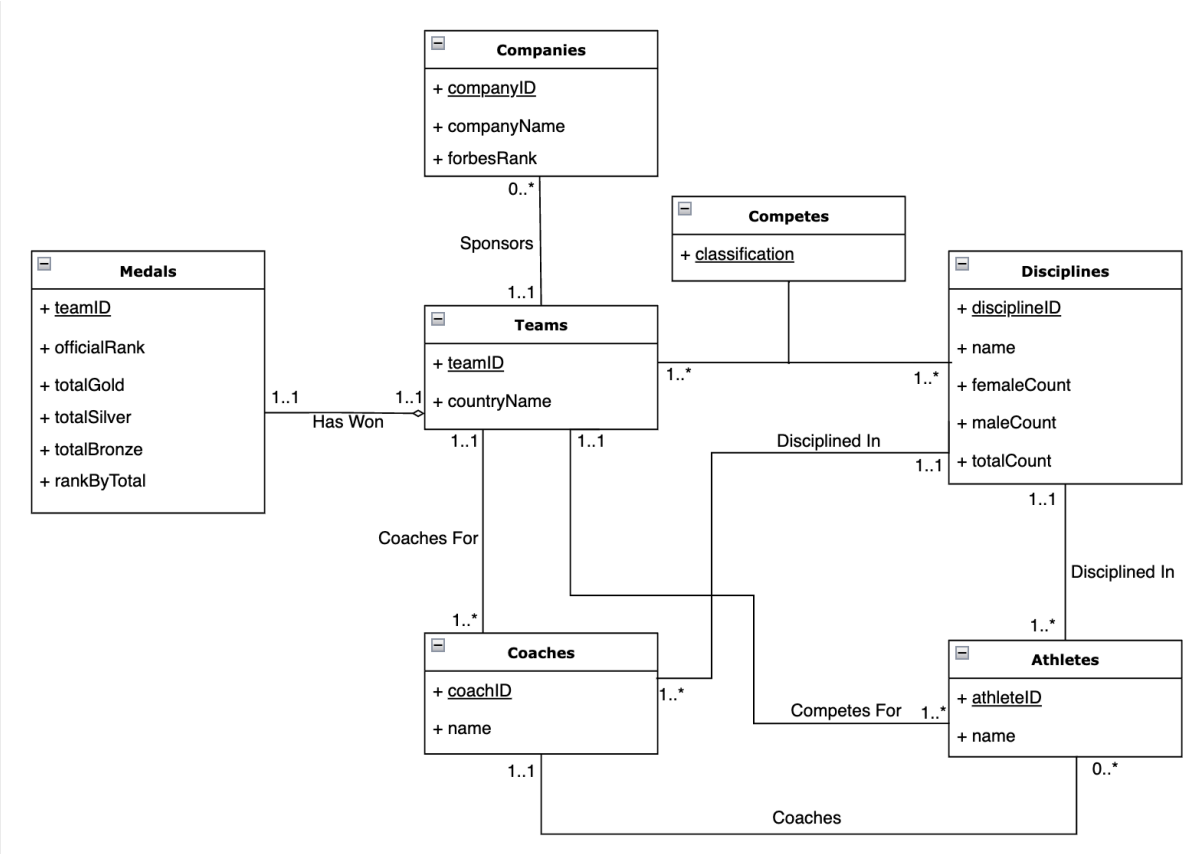
UML Diagram:



Relationship Breakdown:

Coaches - Any coach can have zero or more athletes that they coach which are classified by coachID and athleteID. Each athlete must have the same discipline and country as the coach in this relationship (Because of missing data many new coaches were generated with random names)

Competes - Any team can compete in one or more disciplines, which are classified by the subevent (ex: USA competes in Volleyball, Mens and Volleyball, Womens). Further, any discipline can have one or more teams that compete in it.

Disciplined In (Athletes & Disciplines) - A specific athlete can compete in exactly one discipline, this does not account for subevents. A specific discipline can have one or more athletes that compete in it.

Disciplined In (Coaches & Disciplines) - A specific coach manages a team in exactly one discipline, this does not account for subevents. A specific discipline can have one or more coaches that coach a team for that discipline.

Competes For - A specific athlete competes for exactly one team (i.e. Michael Phelps swims for team USA). A specific team (country) can have one or more athletes that compete for it.

Coaches For - A specific coach coaches a group of athletes for exactly one team. A specific team (country) can have one or more coaches that coach for it.

Has Won - Each team has won a certain number of gold, silver, and bronze medals and achieved a certain ranking. This is represented by a weak entity relationship

Companies - Each team can be sponsored by 0 or more companies which can be identified by the PKs. Each sponsor can sponsor exactly one team, which makes this a one to many relationship.

For our datasets, we make several assumptions, which include the following:

- Every team has at least one coach
- Every team has at least one athlete that participates in it
- Every discipline has at least one athlete that is competing in it
- Every discipline has at least one coach
- Every sport has at least one team that competes in it
- Every team competes in at least one sport
- A coach can have zero athletes that they coach
- Each company sponsors exactly one team

Decomposition

For decomposing our relations, we chose to follow the BCNF decomposition process. By definition this means that for each nontrivial functional dependency, each LHS is a superkey. We chose this format because it is more strict than the 3NF but usually involves less decompositions into smaller tables. Additionally, BCNF also avoids information loss and reduces certain types of redundancies which made it ideal as a choice for our project.

For our relations most of the dependencies have the primary key on the LHS, which means that they were already in BCNF. For the other dependencies, each FD still incorporates some type of super key (not necessarily the candidate key) on the LHS.

Normalizing

The left-hand side of each dependency is a superkey, and there are no partial dependencies so the relations are BCNF.

Companies(companyID, teamID, companyName, forbesRank)
Functional Dependencies: companyID -> companyID, forbesRank, teamID

Disciplines(disciplineID, name, femaleCount, maleCount, totalCount)
Functional Dependencies: disciplineID -> name, femaleCount, maleCount, totalCount

Competes(teamID, disciplineID, classification)
Functional Dependencies: teamID, disciplineID -> classification

Teams(teamID, countryName)
Functional Dependencies: teamID -> countryName

Medals(teamID, officialRank, totalGold, totalSilver, totalBronze, rankByTotal)
Functional Dependencies: teamID -> officialRank, totalGold, totalSilver, totalBronze, rankByTotal

Coaches(coachID, name, disciplineID, teamID)
Functional Dependencies: coachID -> name; disciplineID, teamID

Athletes(athleteID, name, disciplineID, teamID, coachID)
Functional Dependencies: athleteID -> name; disciplineID; teamID -> coachID

Relational Schema

1. Teams(teamID: INT [PK], countryName: VARCHAR(225))
2. Medals(teamID: INT [PK] [FK to Teams.teamID], officialRank: INT, totalGold: INT, totalSilver: INT, totalBronze: INT, rankByTotal: INT)
3. Disciplines(disciplineID: INT [PK], name: VARCHAR(225), femaleCount: INT, maleCount: INT, totalCount: INT)

4. Coaches(coachID: INT [PK], name: VARCHAR(225), disciplineId: INT [FK to Disciplines.disciplineID], teamID: INT [FK to Teams.teamID])
5. Athletes(athleteID: INT [PK], name: VARCHAR(225), teamID: INT [FK to Teams.teamID], disciplineId: INT [FK to Disciplines.disciplineID], coachID: INT [FK to Coaches.coachID])
6. Competes(teamID: INT [PK] [FK to Teams.teamID], disciplineID: INT [PK] [FK to Disciplines.disciplineID], classification: [PK] VARCHAR(225)))
7. Companies (companyID: INT [PK], name: VARCHAR(225), forbesRank : INT, teamID: INT [FK to Teams.teamID])