Part 1

We must model a database in which we are going to store information related to vaccines. The main idea is that any researcher can include a vaccine that has been found against a pathogen and that this vaccine is visible to everyone (Open data). In this database will be stored pathogens, vaccines, vaccines related to genes that will include genetically modified pathogens and carrier responses, including specific responses from the carrier genes.

The table in which the pathogens are stored includes general information about microbial pathogenesis, the range of both natural and experimental hosts, the carrier's immunity, the name of the pathogen, the name of the disease, the date the pathogen was entered into the database, the name of the introducing researcher, and a description of the pathogen.

On the other hand we have vaccines. A vaccine is associated with a maximum of one pathogen, but a pathogen can be associated with more than one vaccine, although there may not be a vaccine associated to a pathogen. Each vaccine has a unique identifier. In addition, the type of vaccine, method of preparation, recommended storage conditions, gene manipulation, name and description are stored.

Vaccines are usually tested on multiple animals using different vaccination protocols, called hosts. Hosts have an identifier, the name by which they are normally known, and the scientific name. During the tests, the physiological and immunological responses of the host are recorded, having into account that a test is always associated with a host and that a vaccine can be associated with many tests. For each test, the host strain, vaccination protocol, persistence, side effects, host immune response, refutation protocol, observed efficacy and a description of the test performed are stored. This information is essential for understanding and comparing different vaccine development strategies.

To specify and summarize the response of a host in each test, it is necessary to store the information that allows summarizing how a vaccine has affected a specific gene of a given host. This information is always associated to a test, while a test can be associated with many gene's response, although it may be not associated with any.

Information related to genetic engineering applied to the genes of the pathogens that have been used in the development of each vaccine is also stored. In this case, the type of engineering applied, the description and the name of the process will be saved. This "engineering" is always associated with a vaccine, and only one type of engineering can be applied to a vaccine. A vaccine could have not been modified.

Finally, we need to save information regarding the genes that have been studied. It is necessary to save the name of the gene, the identifier of the gene according to the NCBI, the DNA sequence to which it belongs, the name of the protein, where the gene begins and ends, the protein weight and the protein length. These genes can be associated with several responses from different genes in different hosts, and a response from one gene in a host is always

associated with a gene. On the other hand, the gene is also always associated with one type of engineering, and one type of engineering is always associated with a gene.

Part 2

Transform the following entity-relationship diagram into the relational model. Do not improve the entity-relationship diagram or assume that there are cardinalities that would be incorrect.