Business Rules

- A person can collect the data about one or more proteins, but the data of a protein can only be collected by one person.
- A target disease can involve one or more proteins. A protein may be involved in more than one target disease.
- Each protein is tracked using UniProtKB entry.
- Each protein has only one recommended protein name, but can have several alternative names.
- A protein can have several structures, but one structure can only belong to one protein.
- A protein can have several domains, and a domain can belong to several proteins.
- A protein can only be encoded by one gene, but a gene may encode multiple different proteins.
- Each protein must have a unique UniProtKB entry, a unique URL to access HTML page from the UniProt database, a protein name, some protein attributes like protein sequence length and protein mass.
- A gene encoding a protein has only one UniProt gene name, one approved HGNC symbol and one approved HGNC name (full gene name), but can have several alternative HGNC symbols and names.
- A gene's UniProt gene name can be different from its approved HGNC symbol.
- Each gene encoding a protein must have a unique HGNC ID, a unique URL to access HTML page from the HGNC database, a UniProt gene name, an approved HGNC symbol, an approved HGNC name and a Chromosomal location.
- A protein can be involved in different subtypes of a target disease. Except the target diseases, a protein can also be involved in other types of diseases.
- Different proteins can be involved in the same disease.
- Each disease must have a UniProt disease name and a MIM number. Different diseases in the UniProt database may have the same MIM number.
- A MIM number can correspond to 0,1 or more DO terms. A DO term can correspond to one or more MIM numbers.
- A UniProt disease name can correspond to 0,1 or more DO terms. A DO term can correspond to one or more UniProt Disease names.
- A DO term must have a unique DOID and a name.