**KW Partner:** UPM
IB Member: Robotiker

1. General Description of Business Use Case

To provide an unified point of access to different biological data repositories: accessible through internet (Nucleotide Sequences, amino acid sequences,…), corporate databases, results of experiments (DNA-chips), health cards, medical literature sites…

Generation and extraction of knowledge from biological data by means of ontologies, combining them (ontology merging) and/or associating them (ontology mapping) to be exploited by means of annotations, intelligent agents, semantic web services and/or semantic grid.

Currently, a great diversity of biological data repositories exists: databases accessible through Internet, corporate databases and microarrays experiments results among others. Equally exists a great diversity of ontologies for model this data. Therefore the situation that the researchers has to face with is a lot of disperse data and different disconnected and poor friendly tools to access such data, therefore the researches have to confront great difficulties to aggregate all the data to carry out the research tasks in an integrated way.

Up today ontologies in biology were considered as mere guides for data structure, with the only purpose to access to the more adequate documents and articles to the researcher interests. This new vision will allow, combining and associating existing ontologies in the biological field, an integral modelling of the biological data sources (genomics, proteomics, metabolomics and systems biology). Once modelled, the annotations, intelligent agents, semantic web agents and the semantic grid will offer a centralised access point to extract and generate knowledge from the biological data repositories.
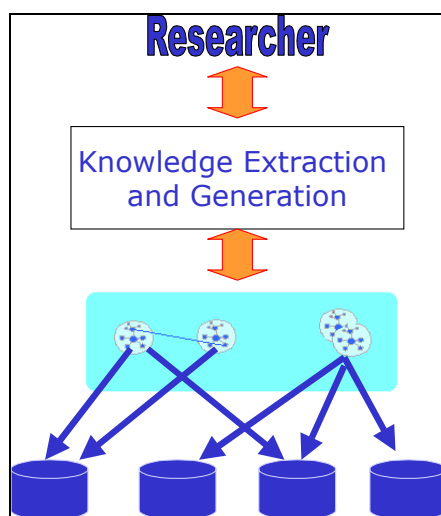
2. Proposed Semantic Web-based Solution

Application of semantic technologies to solve the inherent features of the biology field: huge quantity of disperse-distributed- and autonomous data with great difficulties to be integrated due to their differences in: Terminology (synonyms, aliases …), Syntax (spelling, file structure …) and Semantics (intra-/interdisciplinary homonyms.

Ontologies describe the vocabulary of the data stored at each repository, Annotations describe the data stored at each repository and link it with the corresponding ontology. Ontologies merging and mapping techniques allow integrating these repositories in a consistent and unified way.

3. Identified Research Challenges

3.1 Ontology mapping and merging (Heterogeneity)

3.1.1 Problem Statement

Ontology Merging allows to know which concepts in ontology A are the same than in other ontology B. Detecting common concepts and allowing the "jump" between ontologies. The ontology merging could be used by a company that wants to use de facto standard ontologies associating them to specific company's ontologies. In this way proprietary repositories can be "linked" with public ones.

Ontology Mapping allows to sum up one ontology C with other ontology D to obtain a more complete ontology. Due to the fact that a protein could be implied in cell signalling as in a biological process, summing up two ontologies, one describing cell signalling and other describing biological processes can give us a general overview of a protein function.

Once we have merged or mapped the ontologies at the above task we have to be capable linking these resulting ontologies with public or proprietary data repositories through semantic annotation. Deep annotation is a framework taken into account at this stage.
Note that the owners of the databases have the knowledge, so they are the ones that should annotate the databases with the appropriate ontology.

3.1.2 Knowledge Processing Task and Component

3.1.3 Requirements Analysis

3.2 Knowledge extraction from Ontology annotated content

3.2.1 Problem Statement

When the biological data repositories are linked to the ontologies by annotation, now we have to offer to the researchers' tools to extract knowledge from these sources. In this task existing semantic and artificial intelligence technologies applicability and necessities detection will be carried out.

Up today there are several Data Mining tools focused to solve the immediate and concrete problems that the researchers in the biological field faced day by day. With the current tools it is still needed to understand the structure of the different databases, as with SRS, so the

inherent problems of terminology, syntax and semantics are still present. It is different to know if a table in one database called Organisms is the same table called Species in another database. With TAMBIS it is necessary to define a wrapper service for each database to translate the queries so no automation is provided. To design the wrapper it is necessary to study the database before including it into the system.

3.2.2 Knowledge Processing Task and Component

3.2.3 Requirements Analysis