Natalia García Sánchez and Ana Solbas Casajús

ONTOLOGY CONCEPTUALIZATION – ASSIGNMENT 1

Ontology creation involves a series of subsequent iterative steps for correct modelling. In this initial stage of the process of ontology creation is the ORSD (Ontology Requirements Specification Document), where one should decide what the ontology **final objective** is, the **concepts** it will cover and how they will be **applied** through defining **domain** and **scope** of the ontology, all while specifying **how detailed** the ontology will be depending on the project configuration. For this step it is critical to generate an approximate outlook of the problem through expert knowledge of the fields involved in it.

Ontology Requirements Specification Document	
1	**Purpose**
	In the present work we propose the ontology **Open Bacterial Ontology for Infectious diseases, O(h)BOI!**, that will create a knowledge model base around human bacteria-caused infectious diseases comprising medical, epidemiological, microbiological and molecular data. To this end, this ontology is intended to be useful as means of **surveillance** in case of the rise of an **epidemic**. Having epidemiological data linked to molecular data present in pathogens and underlying disease pathogenesis will help fasten tracing data and predicting patterns in the progression of an epidemic or a specific disease. An **initiative of bacterial emergence control** through computational approaches is decisive and urgent, especially in the current state of affairs where bacterial resistance to antibiotics threatens to cause a global crisis across public health, economy, and ecology.
2	**Scope**
	The scope can be defined by functional requirements, in turn determined by the Competency Questions (CQ) through SPARQL queries that we expect our ontology to answer in the knowledge domains of our ontology. This ontology represents information from microbial phenotypes and metabolism associated to bacterial pathogens, as well as pathogenesis mechanisms, transmission and resistance status to treatments. It also represents approved treatment practices and symptomatology of infectious diseases along with the pathways that are affected in it. Some markers and chemical information about drugs and toxins combined with epidemiological history/ databases/clinical assays and geographical knowledge.
3	**Implementation Language (optional)**
	Ontology Web Language RDF-XML
4	**Intended End-Users (optional)**
	Epidemiological experts in Public Health Organizations Physicians (internists, general medicine practicians, public health specialists…) Scientists such as microbiologists and biochemists for research uses Bioinformaticians Pharmacologists
5	**Intended Uses**

	Decision making in infectious emergent disease control strategy, plan, and action specification. Basic medical practice decision making support. Research purposes in microbiology, epidemiology or pharmacology.
6	**Ontology Requirements**
	a. Non-Functional Requirements
	• Ontology should be written in English • It will have a free open-access license
	b. Functional Requirements: Lists or tables of requirements written as Competency Questions and sentences
	Competency Questions: • What pathogens cause Y infection? • What treatment (if existing) is approved for X infectious disease in Z country? • Is there any antitoxin for a specific bacteria toxin? • Is there any antibiotic to treat X disease? What is the antibiotic mechanism? • How is the disease X transmitted? • What is the prevalence of for X disease in Z country at a specific U time? • In which city C were the first cases of a disease X reported? • Which organs and pathways are affected by a disease? • What is the phenotype P of the organism that is causing the disease? • Which symptoms S are associated with a disease X? **Natural Language Requirements**: • A city **belongs to** a country. • A country can have a border **with** countries. It can also lack borders with other countries (in the case of island-like isolated countries). • If a country X has a border with a country Y, country Y will also have a border with X. It is a **symmetric** property. The range and domain of the property hasBordersWith are a country. • A country can **belong to one or more** continents (e.g., France belongs to Europe and South America as it has French Guiana). • **Belong to** is not a transitive property as a city will not always belong to all the continents that its country belongs to (e.g, Paris belongs to France and Europe, but not to South America). • An antibiotic can be commercialized **by** companies, which can have Headquarters in a Country. • Infectious Diseases are types of Diseases. • An Infectious Disease can have **one or more** Bacterial Pathogens. They can also be tackled through ordering Medical Practices or Prescriptions • Infectious Diseases can also have affected organs, cause symptoms, and affect pathways. A Drug can be directed as a Prescription for an Infectious Dsease. • An Infectious Disease can be transmitted through Vertical Transmission (*e.g.,* mother to child), Vectors (*e.g.,* mosquitoes), or Contact Transmission (when there is physical contact between an infected person and a susceptible person) • An infectious disease can **affect** an organ(s) and/or a pathway(s). • An organ can be a part of (a) system(s) (*isPartOf)* • Bacterial Pathogens are **equivalent to** bacteria that cause at least one disease and

	are a Pathogen subclass. • Pathogen and a non-pathogen are **disjoint** classes. • A non-pathogen **is the complement of** a pathogen. • Bacteria have **phenotypes** (metabolism, staining and morphology, subclasses of phenotype). • All bacteria that are not Gram + or Gram - stained, belong to the class **Gram variable**. • A Gram variable staining bacterial phenotype **is equivalent to** a Gram indetermined staining bacterial phenotype. • Gram +, Gram variable and Gram + staining classes are **disjoint** classes, their **union** conforms the Staining class (exhaustive disjoint). • A coccobacilli class **is the intersection between** cocci **and** bacilli morphology cell shape classes. • Bacteria that are not cocci or bacilli have a irregular shape. They are **disjoint** classes. • A bacteria can belong in classes of Aerobic oxygen requirements, or in classes of Anaerobic oxygen requirements. The union of the Anaerobic and the Aerobic classes conforms the Oxygen Requirement class. • Bacteria can belong to different Assimilation Metabolism classes. • A bacteria can belong in classes of Autotrophic Processers for Energy Sources, or in classes of Heterotrophic Processers for Energy Sources. The union of the Anaerobic and the Aerobic classes conforms the Energy Source class. • The union of Prescription (e.g. drugs) and Medical Practice (e.g. protocols, medical procedures, interventions…) conforms the Treatment class. • Drugs have a brand name and common name. • Drugs have Mechanisms that can affect Targets. • All Drugs are Substances. • Drugs can have Adverse Reactions. • Substances have a ChEMBL ID that identifies them in the database, and an ATCClassification code that identify their therapeutic and chemical properties. They are associated with Chemical Entities as well. • Both **antibiotics** and **antitoxins** are drugs. • Chemical entities are small molecules that have a **chemical formula**, a **canonical smiles** code that identifies them and a molecular mass **weight**. • Bacteria can have **genes** that encode **proteins** that can be **toxins**. However, not all toxins are proteins. • Toxins can be **endotoxins** or **exotoxins** (disjoint classes) depending of whether they are secreted from the bacteria or not, and their **union** conforms the toxin class (exhaustive disjoint). • **Enterotoxins** are toxins that affect the intestine or the stomach. • All genes have a **symbol** and a location in the genome (string as it can contain the position and the chromosome/plasmid) • The **length** of a protein is the number of amino acids of its primary sequence. Proteins can have associated names identifying them. • Toxins have **targets** that can participate in the host metabolic **pathways**. • **Antitoxins** are designed to neutralize toxins. • **Antibiotics** have a mechanism of action associated.
7	**Pre-Glossary of Terms (optional)**
	a. Terms from Competency Questions

	Infectious Disease, Treatment, Bacteria, Bacterial Pathogen, Pathogen, Non-pathogen, Infectious agent, Gram +, Gram variant, Gram indetermined, Gram -, Drugs, Medical Practice, Bacterial Phenotype, Antibiotic, Anti-toxin, Prevalence, City, Country, Continent, Pathogenesis mechanism, Symptoms, Organ, Pathway…
	b. Terms from Answers
	Treatment, Pathogen, Bacterial Phenotype, Antibiotic, Anti-toxin, Prevalence, City, Country, Symptoms, Organ, Pathway…
	c. Objects
	hasAffected, hasBordersWith, hasApproved, hasManifestationOf, isManifestedIn, hasPhenotype, hasMorphology, hasMetabolism, Causes, IsCausedBy, hasCapital, isCapitalOf, belongsTo, isLocatedIn, hasFirstCaseReported...