Assumptions

• There is value in alignment with OBO/BIOTOP
  • OBO - http://www.obofoundry.org/
  • BIOTOP - http://www.obofoundry.org/
• The organism concepts describe organism populations not biological taxa.
• The model will depend on class definitions that are intrinsic to bacteria
  • Shapes, staining characteristics, growth requirements
  • Biologic functions (e.g. sugar fermentation)
  • Living organisms participate in processes
  • Proper physical parts (material entities) related however distantly to (human) anatomy.

The Open Biomedical Ontologies http://www.obofoundry.org/ exist as a foundation improving sharing amongst useful ontologies (e.g. Gene Ontology GO). As I understand Biotop organism classification, it is based on connection between the organism cell and a (primitive only) hierarchy of “Taxon qualities.” I’m not fond of this for a couple of different reasons:

1) The major reason for an organism (re-)model is to facilitate maintenance which we think will be improved through meaningful autoclassification.

2) Philosophically (pun intended) I think the Taxon’s are manifestations of the characteristics of actual organisms (and classes) and not the other way around (a point of distinction with the Biotop view). Taxon quality is useful (eventually anyway) for content like “Human insulin produced by E. coli” but I don’t like the idea that it drives the organism model. I suspect that BIOTOP has done this because the actual differentia for eukaryotes are (going to be) incredibly difficult to identify and model.

The microbiology laboratory reports that I see are presented (something like): “Isolate: E. coli” The literal meaning of this report is that a member of or a population from the species E. coli is present. It does not mean that the entire E. coli taxon is in the patient (so the SNOMED concept is NOT “E. coli species” it is “member of E. coli species”. It has been our intention all along to make the organism hierarchy as close to
a realist ontology as we can within SNOMED CT’s constraints. We believe that confining the definitions to intrinsic qualities of bacteria allows the definitions to align with the actual laboratory procedures performed in identification. Extrinsic characteristics (e.g. causes diarrhea, is pathogen etc.) are likely patient state dependent (immune function, diet, species, etc.)
Access this TOC in “slide show” mode:

- Model diagrams (12 slides)
- Quality Values (8 slides)
- Authorities (4 slides)
- Naming convention (1 slide)
- Concerns and limitations (4 slides)

Return to TOC
I think this is sound. A set of organism qualities in support of bacteria only is likely very small. Recommend using or creating content from Quality values hierarchy.
I like the grouping of “absolute nutritional requirement” with (a potentially long list of) substances. I dislike the idea of “absolute nutritional requirement” as either function or process. Treat as a Quality Value. There are potentially two views of this representation. The cysteine requirement is either “absolute nutritional requirement for cysteine” or “absolute inability to produce cysteine”
IF there is value in aligning with BIOTOP (I believe that the value is not immediate), this is one places that this model must make some arbitrary decisions about just how tight the alignment might be. To my eye, BIOTOP’s assignment of “Function” and “Process” as well as the domain ranges of attributes make more sense for multi-cellular eukaryotes but may not have direct applicability to prokaryotes. Recommendation from Dr. Case and I is that for prokaryotes at least, we should prefer functions and the attribute “bearer-of.”

Based on BIOTOP’s Domain restrictions, these are the attribute/value pairs:

**Attribute: BIOTOP “bearer of”** – inverse of BIOTOP “inheres-in” - definition "inheresIn (inverse: bearerOf) relates a quality, role, function, disposition, or information object with the physical entity it depends on. (Note that for processes there is a separate relation pair "hasProcessQuality" and "processQualityOf") Example: a color inheres in a paint, the ability to fly inheres in a bird, or a pdf file inheres in a USB stick."

**Value Domain: BIOTOP “Function”** – definition "A classical definition of function according to Wright is that the function F of X is Z means that X is there because it Zs, and Z is a consequence of X being there. For artefacts, functions are distinguished from dispositions by the purpose they have been built. For example, a hammer has the function to drive in nails, but not to be used as a weapon. However, it has the disposition to be used as a weapon under certain circumstances. For biological
objects, which developed by evolution, the definition of function is still subject to controversy (e.g. Barry Smith's view of function as pertaining to a canonical life plan, cf. http://www.slideserve.com/presentation/103450/The-Canonical-Life). Note: In BFO "Disposition" and "Function" are siblings, in BioTop Function is more specific.

**Attribute: BIOTOP “Process”** – definition "Process is in BioTop the generic subsumer of anything that "occurs". Processes can span across time and have temporal parts (i.e. there is no time in which all parts of a process are simultaneously present). A special kind of process is the event, which has no temporal extension. It can be regarded as Process Boundary (BFO). Processes have physical or informational entities as participants."

**Value Domain: BIOTOP “participates in”** - definition "HasParticipant (inverse: participatesIn) relates a process with a non processual entity which plays some role in the process. Process participants may exist during the whole process, remain unchanged or undergo changes; they may come into being or get out of being during the process. Process participation is distinguished from process location. Example: An urea molecule which is excreted in a renal filtration process is participant of this process. A person who undergoes an operation is participant of this process."

5
Aerobic respiration can certainly be thought of as a process and from the perspective of an oxygen molecule it absolutely would be. It’s also a BIOTOP process within a multi-cellular organism (requires participation of lungs and blood stream and mitochondria, etc.) On the other hand it is an intrinsic activity conducted throughout the life of an aerobic bacteria. We believe that for prokaryotes at least, respiration can reasonably be thought of as a function rather than a process. It may or may not be necessary to specify “bacterial aerobic respiration” as a special case...
This is another case where “bearer of function” may not be “BIOTOP correct.” On the other hand, antimicrobial resistance might simply be a quality and this construction would be proper.

In the second example, with a vague term like “multi-” there is no way you can be more specific than “antimicrobial agent” as the value for “towards.” The question mark is really about likely intended meaning of the concept. The concept “Antimicrobial resistant E. coli” could be modeled the same as Methicillin resistant with the class “Antimicrobial” in place of the class “Methicillin” Taken literally, the concept means resistant to one or more antimicrobials, then those same pairs would have to be augmented with a data property that indicates “more than one of.” Precedent for this approach is established for pharmaceutical products and ingredients.
Bacteria structure model
Gram negative bacteria

Return to TOC
Bacterial antigen is the SNOMED class Bacterial antigen (substance) 116634000. In a generalization of the model, it would properly be labeled “some bacterial antigen.”
SNOMED editorial policy requires individual role groups for each antigen (Two more role groups will be added to this definition).

Salmonella Nitra is a name assigned by Kauffman White to represent this seroType. This class exists in nature. At present, Salmonella Nitra (organism) and Salmonella Nitra are FSN and PT respectively.

The next example (S. Typhimurium) is a serotype label created by Kauffman White to represent 4 extant Salmonella serotypes.
This is an example of a class that is really an “administrative abstraction.” The serotype does not exist in nature but represents the classes that are optionally included under the serotype name “Salmonella Typhimurium.”

Salmonella I 4,12:i:1,2 ; Salmonella I 4,5,12:i:1,2 ; represent an actual isolates that KW says should be named Salmonella Typhimurium. Antigen 1 is conditional (appears with appropriate phage conversion). Antigen 5 is optional (may be present or absent without changing the KW class).

We have not determined means by which the optionality of the various antigens can be managed using the SNOMED DL. It’s not even clear THAT they need to be represented in anywhere other than concept names.
“Morphology is a branch of biology dealing with the study of the form and structure of organisms and their specific structural features.”

These qualities are the “visual morphologies” of prokaryotic cells. SNOMED’s morphology hierarchy is really populated with “morphologically abnormal structures (of patients)” As these are simple biological shapes, the morphology hierarchy of SNOMED seems inappropriate. It seems appropriate to place these as subtypes of either “Shape finding (qualifier value) 107644003 or “Formations (qualifier value) 255277001“ (as a sibling to Shape finding).
Use BEARER-OF these qualities.
This is a relatively weak part of the model. Not clear

Growth requirement innate to the organism + the specific substance required. Role grouping is proposed to link the category (e.g., “absolute requirement for growth”) to the substance.
Flagella and Capsules are proper physical parts of prokaryotes.

Proposal: Create “Prokaryotic cell structure (cell structure).” Subsumes existing cell wall concepts. The desire is to represent the anatomical structures necessary to represent prokaryotic organisms. Various patterns of flagellar distribution exist. These are not truly subtypes of flagella. This is a list of structure distribution, not of structures per se. Determine the value of making this distinction. This actually may only be required for proper classification of single cell Eukaryotes. Phase 2 analysis? Maybe a value in a role group with Has-PROPER-PHYSICAL-PART = Flagellum. Keep these in the body structure hierarchy with top level concept of “prokaryotic cell structure” and immediate children of “prokaryotic part” and “prokaryotic morphology” Use the same structure as “Shape”. These are qualifier values for now. Handle as structure? Need more background on this. Smooth and rough are probably not useful. May be colloquial synonyms.
The gram stain test determines whether or not a bacteria has a cell wall that can retain crystal violet when a particular staining technique is followed. As an organism characteristic, the test determines whether the cell wall of the bacteria is of a gram variable species include Genus Mycobacterium which are classified as acid-fast gram positive but also as gram-variable based on the staining results. Essentially they have a gram positive cell wall but lack an outer membrane.
Where do the functions live? No way to represent inherent functions at this time (vs. observation of function values). Facultative anaerobes have the disposition to be anaerobic under specific conditions, so is really a subtype of aerobe. Is this a function? Process? Change to Respiration type? A Conditional state?

For classification purposes, it is not likely necessary to include the absence of a biochemical process/function.
These are complex interactions between bacteria and specific media. Appropriate attribute is not certain. Motility is tricky from the laboratory perspective. In serotyping at least, “non-motile” is often substituted as an interpretation for “no flagellar antigen detected” (We think this is inappropriate). What is the specific meaning in a laboratory? (We think it should be that motility has been demonstrated). Is motility even required for differentiation (we will learn).
Serotyping is (currently) based on immunologic detection of antigens that are either embedded in cell surface structures (e.g., outer membrane) or are components of other bacteria “parts” (e.g., flagellar antigens). General antigen classes such as Salmonella O (somatic) antigen (substance) 103109003 exist in SNOMED. Specific antigen substances do not. Q: Does staying in the substance hierarchy facilitate connections to procedures?

Antimicrobial resistance is similar in that it is the expression of a quality that applies to a range of substances (antimicrobials).

In both cases role grouping is used to associate the specific substance with the quality (resistance). Is there a use case for expressing both resistance and susceptibility? Existing general classes (e.g. Methicillin resistant Staphylococcus) are based on resistance and that is the primary clinical concern.
Organism (species) subtypes often include groupings based on some description of pathogenic mechanism. This list reflects the existence of three distinct kinds of “Biotype.” In the Yersinia example, the labels “pathogenic” and “non-pathogenic” appear to be about the interaction between the bacterium and the host. In this case however, the correlation between certain intrinsic characteristics and the extrinsic activity is strong. Toxin production seems (to JRW) to be a characteristic of the organism and could be modeled much as certain biochemistries (e.g. catalase production) will be modeled. Diarrheagenic and caseating are really findings in an infected patient not intrinsic properties of the organisms.

JRW and JTC have not reached full agreement on this point.
Locations for values

• Values for “bearer of” role
  • Shapes (e.g. round) are qualifier values, is coccus-shaped just a specialization of round?
  • Nutritional requirements?
  • Functions and processes?

• Values for “proper physical part” role
  • Currently cell structures in “Anatomical structure” hierarchy
  • Suggest a “prokaryotic cell structure” grouping

• Values for “component part” role
  • Current proposal is substances
  • Current “bacterial antigen” placement is questionable
    • Bacterial antigens are not thought of (first) as allergens.
Authorities

• Bacterial nomenclature:
  • Very mature systems for identifying and maintaining
  • Probably why Bacteria hierarchy in SNOMED is as up-to-date as it is
    • LPSN (http://www.bacterio.net/) - aligned with ISEM.
    • DSMZ-Prokaryotic Nomenclature Up-to-date (https://www.dsmz.de/bacterial-diversity/prokaryotic-nomenclature-up-to-date.html) -- used by the WHO for WHONET

List of Prokaryotic Names with standing (LPSN) Site is in alignment and follows the ‘International Committee on Systematics of Prokaryotes’ (ICSP, http://www.the-icsp.org/) recommendation as published in the International Journal of Systematic and Evolutionary Microbiology (http://ijs.sgmjournals.org/).

DSMZ IS current and up-to-date but also includes out-of-date names. Inconsistent representation of levels below species. Non-standard labels (names) for existing serotypes.
Authorities

• Viruses
  • CONSIDERABLE IMPROVEMENT in taxonomic alignment (with other organisms) and published maintenance principles
  • Would rate this as reliable.
  • available Virus taxonomy resource is greatly improved
    • http://www.ictvonline.org/virusTaxonomy.asp
Authorities

• Fungus
  • Detailed analysis has not been done at this time.
    • http://www.mycobank.org/
    • www.doctorfungus.org (used for WHONET)
Parasite is not a proper classification for an organism except in the context of the parasite-host relationship.

A grand remodel of organisms will require authorities on the actual Linnaean classes that participate in parasitism.
Issues...

LPSN does not include taxonomic rank in names EXCEPT at the subspecies level. The reason for the inclusion of “subsp.” in each bacterial subspecies is unclear to us. Our design principles to date have included the idea that we will NOT include taxonomic rank in FSN’s. Certain users would like an exception made to our design principles for the Genus level. Believe this is to facilitate aggregation in queries such as “count of all (subtype) isolates of the gammaproteobacteria grouped by genus”
Concerns – Complexity and Uncertainty

• To some extent (especially among enterobacteriaceae), bacterial identification is probability based. (e.g., 50% of E. coli are sucrose fermenters). See:
  • Need copy of Bergey’s to evaluate ability to produce solid “always true” differentia for bacteria classes
Concerns – genetic identification

- There is fairly clearly a shift going on towards genetic identification (e.g., PCR) and away identification by classic.
  - Does this mean genetic CLASSIFICATION or does it just mean “shortcut” to the existing phenotypic classification?
    - Facilitate automation and speeds detection/screening methodologies.
  - Impact on phenotypic hierarchy is unclear at this time.
  - Have seen nothing to suggest that the bacteria hierarchy will undergo wholesale change.
Concerns – Groups and complexes

• Unable to identify a reliable reference concerning CDC “biogroups.” SNOMED has 30+ of these already. There seem to be more (SNOMED has 1 under proteobacteria, as many as 11 exist) I believe these to be of two types.
  • For newly discovered / unclassified organisms
    • Many now have names (not clear if CDC still uses or expects use)
      • See Bruckner and Colona
      • CDC Enteric Group 59 = Buttiauxella noackiae?
    • Importance of cross-reference unclear
    • CDC policies on use of these names vs. “official” names are unknown. (requirements to submit?)
  • For organisms that CDC wants submitted before classification is complete.
Concerns - miscellaneous

• Interest has been expressed in an ability to track “previous names / historical names”
  • Unordered synonyms?
  • Concept retirements and referrals
• Concepts with non-specific references to attributes (e.g., Multiple antimicrobial resistant X)