

A semantic index of phenotypic and genotypic data

[Charles Parker](#)¹, Nenad Krdzavac², Kevin Petersen², Amber Roberts², Grace Rodriguez² and George Garrity², (1)NamesforLife, LLC, East Lansing, MI, (2)Microbiology & Molecular Genetics, Michigan State University, East Lansing, MI

While it is generally acknowledged that microbes are an invaluable source of commercially useful new products and processes, discovery is often based on finding not only the right strain but also the right growth conditions to achieve a desired outcome. Access to accumulated data and knowledge about the metabolic and genetic potential of the strains of interest are essential for success. But not all data are of similar quality nor are all data amenable to modern approaches of computational analysis without extensive cleaning, interpretation and normalization. Key among these are phenotypic data, which are more complex than sequence data, occur in a wide variety of forms, use complex and non-uniform descriptors and are scattered about the literature and specialized databases. To address this need, we are constructing a Semantic Index of Phenotypic and Genotypic Data, built on an ontology of microbial phenotypes and growth conditions extracted from the taxonomic literature. In this project, we utilize a new class of Data as a Service (DaaS), based on reasoning over curated data. This approach provides a powerful query method that is tolerant of missing or ambiguous information. When combined with Description Logics (DL; a family of formal knowledge representation languages) it has the potential to unlock a wealth of hidden knowledge buried in observational data by supporting inferences about properties of interest. The resulting resource will serve as a useful bridge from discovery to production.