

The BioHub Knowledge Base: Ontology and Repository for Sustainable Biosourcing

Warren Read², George Demetriou¹, Goran Nenadic¹, Noel Ruddock²,
Robert Stevens¹ and Jerry Winter²

¹University of Manchester, ²UniLever

Abstract. The motivation of the BioHub project is to create an Integrated Knowledge Management System (IKMS) that will enable chemical engineers to source ingredients from bio-renewables, rather than from non-sustainable sources such as fossil oil and its derivatives. The BioHub Knowledge Base (BioHubKB) is the data repository of the IKMS; it employs Semantic Web technologies, especially OWL, to host data about chemical transformations, bio-renewable feedstocks, co-product streams and their chemical components. Access to this knowledge base is provided to other modules within the IKMS through a set of RESTful web services, driven by SPARQL queries to a Sesame back-end. The BioHubKB re-uses several bio-ontologies to augment its knowledge organisation schema. Parts of plants form feedstocks, while various processes generate co-product streams that contain certain chemicals. Both chemicals and transformations are associated with certain qualities, which the BioHubKB also attempts to capture. Of immediate commercial and industrial importance is to estimate the cost of particular sets of chemical transformations (leading to candidate surfactants) performed in sequence, and these costs too are captured. Data are sourced from companies' internal knowledge and document stores, and from the publicly available literature. Both text analytics and manual curation play their part in populating the ontology. We describe the IKMS, the BioHubKB and the services that it supports for the IKMS.