# RDF Friendly Chemical Taxonomies for Semantic Web

Bhat, T.Narayana[1*], Barkley, John[2]
[1]Biochemical Science Division (831), NIST, Gaithersburg, MD 20899 USA
[2]Software Diagnostics and Conformance Testing Division (897), NIST, Gaithersburg, MD 20899 USA

Drug design, development, and testing is an organized and coordinated effort involving many scientists coming from different disciplines ranging from software engineering to bench science. Web and chemical databases play a critical role in data warehousing and validation efforts in all aspects of this effort. Current techniques and standards of organizing, specifying, and querying chemical data are far from satisfactory. These standards and techniques do not allow the assignment of semantics (context dependent meaning) to the data. Moreover, available standards for representing chemical structures such as the Protein database format, MOL format, CML do not facilitate easy interoperability of structural information among different contexts and users. RDF coupled to chemical taxonomies promises to fulfill the dream of medicinal chemists by providing a unique environment to specify, to exchange and to integrate structural data in a seamless fashion both in batch and web environment with or without the use of databases. I will describe a use case for a chemical taxonomy (http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase ) for drug discovery in the context of AIDS research. At first, a need analysis for the drug discovery is carried out. Then using these needs, a set of commonly asked questions during the drug discovery process are coined. Using these commonly asked questions, chemical data are evaluated and organized into a taxonomical tree that establishes precise one-to-one relationships (RDF) among all components of the data. The emphasis of the chemical taxonomy is on organizing all critical elements of the data that are used to answer the commonly asked questions in a fashion that enables rapid access in a Web environment using general purpose search engines supported by publicly available database software such as ORACLE or MySQL.

During my talk, I will explain the use of this technique for the drug discovery process for AIDS (http://xpdb.nist.gov/hivsdb/hivsdb.html). (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=16508960&query_hl=3&itool=pubmed_docsum). This Web page enables scientists to query and visualize drugs and drug candidates with emphasis to their interaction with the viral targets such as the HIV protease and HIV reverse transcriptase. Such visualization and analysis may reveal the subtle difference between the binding modes of different drugs with a wild type or drug resistance mutant virus.