

Imprintome Project Description to LOD

Send a little self introduction to the mailing list (include an intro to your project and associated RDF Data Sets where such exist or are planned).

My name is Diana Magalhaes de Oliveira and I am an Associate Professor (and also a Principal Investigator) at the State University of Ceara (UECE - www.uece.br/mestrandonutricao) in Fortaleza, Brazil. My main research interests cover bioinformatics methods to analyze genomics data.

Currently I am starting to work in linking data from a particular dataset of human genes in order to provide a comprehensive collection of all (known and predicted to be) imprinted genes in the human genome, termed imprintome. My group and I have examined, compiled, structured and linked data from around 240 unique genes to use it as a sharing resource for genome and epigenome interrogated studies. A detailed biological description of imprinted genes with their roles and features is out of the scope of this work, but we rather aim at providing a reasonable, valid application of the Semantic Web and Linked Data approaches to a structured dataset regarding the human imprintome. Therefore, we have focused on displaying the big picture of the human imprintome as a large-scale draft map of linked data for enabling the tasks of browsing and mining imprinted genes towards further, domain-expert understanding of their complex nature. Data management operations, specially those enabled by S3QL (<http://s3ql.info>) that can also be formalized in SPARQL (Deus et al., 20110) are planned to be used. There is an attempt to work on the maturation of the core S3DB application and other applications using S3DB's API to interoperate with the S3DB data service (<http://link.s3db.org/owl>), before adopting the final model and via a SPARQL endpoint (approximately 130,000 triples in a N3 file format).