

**Panel:** Innovation on Digital Curation

**Natural language processing to enhance accessibility to knowledge in RegulonDB**

Julio Collado-Vides, Center for Genomic Sciences, UNAM, Mexico

**Abstract**

The ultimate goal of curated databases is to facilitate access to data, information and knowledge. Data and information are usually modeled as objects or tables as part of the structured component of a database. Knowledge can be practically defined as all the additional statements in published journals that because of different reasons (limitations in the database modeling, as well as the rich expression of correlations, and interpretations in written language) are present in original publications but not in the structured items in databases because they are hard to curate and encode. I will present progress on how we have used natural language processing (NLP) to address knowledge encoding in RegulonDB where we have gathered for more than 20 years information on transcriptional regulation and operon organization in *E.coli* K12. We are working on NLP implementations both at the curation front as well as at the users front.

Acknowledgement: This project is supported by NIH-NIGMS grant number RO1GM110597