

ORAL PRESENTATION

Open Access

# ChemSpider - building a foundation for the semantic web by hosting a crowd sourced databasing platform for chemistry

Anthony J Williams<sup>1\*</sup>, Valery Tkachenko<sup>2</sup>, Sergey Golotvin<sup>2</sup>, Richard Kidd<sup>3</sup>, Graham McCann<sup>3</sup>

From 5th German Conference on Cheminformatics: 23. CIC-Workshop  
Goslar, Germany. 8-10 November 2009

There is an increasing availability of free and open access resources for chemists to use on the internet. Coupled with the increasing availability of Open Source software tools we are in the middle of a revolution in data availability and tools to manipulate these data. ChemSpider is a free access website for chemists built with the intention of providing a structure centric community for chemists. It was developed with the intention of aggregating and indexing available sources of chemical structures and their associated information into a single searchable repository and making it available to everybody, at no charge.

There are tens if not hundreds of chemical structure databases such as literature data, chemical vendor catalogs, molecular properties, environmental data, toxicity data, analytical data etc. and no single way to search across them. Despite the fact that there were a large number of databases containing chemical compounds and data available online their inherent quality, accuracy and completeness was lacking in many regards. The intention with ChemSpider was to provide a platform whereby the chemistry community could contribute to cleaning up the data, improving the quality of data online and expanding the information available to include data such as reaction syntheses, analytical data, experimental properties and linking to other valuable resources. It has grown into a resource containing over 21 million unique chemical structures from over 200 data sources.

ChemSpider has enabled real time curation of the data, association of analytical data with chemical structures, real-time deposition of single or batch chemical structures (including with activity data) and transaction-

based predictions of physicochemical data. The social community aspects of the system demonstrate the potential of this approach. Curation of the data continues daily and thousands of edits and depositions by members of the community have dramatically improved the quality of the data relative to other public resources for chemistry.

This presentation will provide an overview of the history of ChemSpider, the present capabilities of the platform and how it can become one of the primary foundations of the semantic web for chemistry. It will also discuss some of the present projects underway since the acquisition of ChemSpider by the Royal Society of Chemistry.

#### Author details

<sup>1</sup>Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, NC-27587, USA. <sup>2</sup>NCBI, NLM, National Institutes of Health, Washington, USA. <sup>3</sup>Royal Society of Chemistry, Cambridge, UK.

Published: 4 May 2010

doi:10.1186/1758-2946-2-S1-O16

**Cite this article as:** Williams et al.: ChemSpider - building a foundation for the semantic web by hosting a crowd sourced databasing platform for chemistry. *Journal of Cheminformatics* 2010 **2**(Suppl 1):O16.

<sup>1</sup>Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, NC-27587, USA

