ISRANALYTICA 2019
The 22nd Annual Meeting
of the Israel Analytical Chemistry Society
CONFERENCE & EXHIBITION
22-23 January,2019, The David Intercontinental Hotel, Tel Aviv, Israel

Evaluating and using big data in chemistry

D. Brynn Hibbert

School of Chemistry,
University of New South Wales (UNSW), Sydney, Australia

b.hibbert@unsw.edu.au

'Big data' means many different things in a subject like chemistry. To an analytical chemist it might be the multi-dimensional output of GC x MS x MS that needs to be calibrated to solve a multivariate problem. To the pharmaceutical or organic chemist integrating NMR with finding targets in a database that might have biological activity is now almost routine. And for the relatively new branch of cheminformatics it might be trawling through clicks on published articles across the world to decide what the trending areas of chemistry are this week.

In fact data becomes 'big' as soon as one person cannot understand them all at one view. If a chemist must resort to a computer to treat the flow of information from an instrument, then the data is already 'big' whether a few kilobytes, megabytes or terabytes.

The International Union of Pure and Applied Chemistry (IUPAC) is "the world authority on … atomic weights and many other critically-evaluated data" (https://iupac.org/what-we-do/). When evaluated data could simply be published in, for example, the latest volume of the Solubility Data Series, keeping up with the flow of information was relatively easy. However the data to be evaluated is pouring in from the vast amount of peer-reviewed (and quasi-peer reviewed) literature, and we need new tools to reduce, visualise and extract useful information from it.

Some problems found in the author's and colleagues' work will be described, including modelling formation constants for the chelation of metal ions by tripeptides (solved), high-throughput screening of catalysts (solved), applying the analogue drug laws (unsolved) and predicting optimum chromatographic conditions (partially solved).