**Big Data in Biomedicine – Challenges and Opportunities for Libraries**

The concept of Big Data refers to data that cannot be handled by simple tools such as spreadsheets or other traditional database formats. The reason for this are three features of Big Data – volume, velocity and variety. Some areas of science, such as particle physics, have been known since long for their huge data output; in biomedicine, the quantity of experimental data has increased exponentially in more recent years, in areas such as genome sequencing, analyses of brain connectivity, or epidemiological and public health studies. Besides quantity, speed of data generation and their diversity have also grown rapidly, owing to technical advances like automated sequencers or to the multi-faceted approaches required by many biomedical experiments.

Traditionally, librarians developed techniques to handle structured information in form of publications – books, articles or other ‘ready-made’ media formats. Nowadays, the policies of many funders require that all so-called ‘raw’ data gained from funded experiments are securely stored and kept available for further use. Besides other novel library services for investigators (such as publication repository management), curation of Big Data is a great opportunity for librarians to apply their mastery of information handling. Big Data mostly come as raw data sets that require adequate ways of preservation and assignment of metadata – subject headings, semantic tags, classifiers and others. These activities connect biomedical data sets to their scientific or clinical context, which in turn makes them retrievable and re-usable.

My talk discusses recent strategies and techniques of libraries and librarians to manage Big Data in a way that maximizes their use for the scientific community.