

The Wheeler Lecture – How Big is My Data

By Professor Chris Budd – Bath University

Professor Budd began his lecture by explaining that personal data are now collected everywhere in vast quantities and these trends are expanding with the development of the "internet of things". As an example of the acceleration of data collection he related the tale of a girl in the US who had a supermarket loyalty card that resulted in her receiving a communication from them stating that as she was pregnant she might like to buy certain products from them. This was to the dismay of her parents who knew nothing of a pregnancy.

He then addressed the aspect of quantitative data with the illustration that the human brain has one Terabyte of capacity, which is considerable when you consider that a proficient Morse code operator could tap out a message at the rate of approximately 2 bytes per minute. As a further example of the capacity of big data he noted that it would take 1,000 years to tap out the data, at 15 words a minute, comprised in a film such as "Bridget Jones' Baby".

Given that the quantity of available data is now so massive the collection and storage is becoming a major challenge with the result that it is only the major private sector digital entities such as Facebook and Google that possess the resources to control data stores that are measured in Terabytes (2 to the power of 50), Exabytes (10 to the power of 18) and even Zetabytes (10 to 21). The collection and handling of such massive volumes of metadata requires the application of linear algebra and eigenvector matrices.

How this data is used is really the forum of the ethicist and the lawyer. It can, like all things, be used for good and for ill. Chris Budd warned the audience to be aware of "leaking data" every time they walked into a shop or clicked the BUY button on their computers.

Since the lecture fell on St.Valentine's Day, Professor Budd discussed briefly the Sultan's Dilemma when choosing a new wife. Supposing that the sultan must select one from a hundred females and cannot return to a previous candidate once he has walked past her he clearly has a problem as the quality of the applicants, in terms of beauty and desirability, is variable. How does he choose? He walks past the first 30. This supplies him with a good representative sample of the quality of the field, from which he can set his standard to judge the best of the remaining candidates. This technique he claimed was equally as valid if using a computer dating agency.

The 2018 Wheeler lecture was extremely apposite in view of the current problems with the misuse of data by Facebook and was very well received by the audience present and as expressed by the number and quality of the questions that followed the talk.