Are we prepared for “Big Data” in high pressure sciences?

***Clemens Prescher**, Institute for Geology and Mineralogy, University of Cologne, 50674 Köln, Germany*

The term “Big Data” is ubiquitously appearing in the media in the last years. Originally the term was chosen for the challenges arising from more and more data available due to cheaper and cheaper information obtaining devices such as e.g. smart phones. However, nowadays the term is more generally used for analysis of existing data and for predictions out of it and therefore obtaining extra value.

While the current amount and type of data in high pressure sciences is not at all comparable to the amount of data what big companies like e.g. Google or Facebook are currently acquiring and have in stock, I think we are still facing similar challenges. The upcoming upgrades to fourth generation synchrotron facilities with MBA lattices, will lead to 100 times increase in brilliance. This not only results in photon hungry experiments becoming more accessible to be performed at high pressure, but also that current standard experiments will become even faster and the amount of data possible to collect increases tremendously. Already today, a complete high pressure single crystal diffraction set can be collected in about 20-30 seconds due to the advance of current detector technology. However, the large rate of data being collected necessitates fast data reduction in order to monitor the physical and chemical processes (phase transition, dissociation, melting, etc.) occurring in the sample and being able to make meaningful adjustments in real time during the course of the experiment. This decision process requires graphical, easy-to-use data reduction and exploration software, for on-the-fly analysis during the beamtime. Otherwise the limiting data analysis capabilities will become a bottleneck during high pressure experiments, and lead to erroneous decisions or waste of beamtime and samples.

Another current issue in high pressure sciences is the availability and accessibility of existing data to the community. While the crystallography community has made large efforts to create a common data format (“cif”) and maintain databases for crystallographic information. The high pressure community is still lacking a useful database for equation of state data for example. This would not only be of tremendous help for new students coming into the field, but also for other communities which require our results like geophysics, planetologists and physicists.

Here, I will present the current status of available software, databases and future opportunities and possibilities for improvement in order for the high-pressure community to become better prepared for “Big Data”.