# Managing large atomic and molecular data sets: HITRAN, ExoMol and CascadesDB

Christian Hill
Atomic and Molecular Data Unit, Nuclear Data Section, IAEA

The rapid improvement in computational power – CPU speeds, primary storage (RAM) and secondary storage (magnetic hard drives and solid state drives) – over the last few years has enabled researchers in many scientific fields to produce large amounts of data in the course of their work. Generally, these data sets must be stored somewhere accessible in a structured way that makes them accessible, searchable and reusable.

In atomic and molecular physics, relevant data sets may comprise collisional cross sections, spectroscopic line lists, energy level lists, and fundamental structural properties. Although currently such data sets would not be described as "big data", they can reach several TB in size and careful planning is required to make them useful and maintainable.

This presentation describes the strategies that may be adopted to curate large atomic and molecular data sets, including the design of relational database schemas, storage models, online interfaces, data interoperability standards, the validation and evaluation of data, data reduction and transformation, and security concerns. Reference will be made to several relevant databases in spectroscopy (HITRAN and ExoMol), plasma physics (QuantemolDB and ALADDIN) and material science (CascadesDB).