

### Session 3: workflow and data management

Running benchmark calculations is tedious. Having a workflow manager in which the procedure to run a benchmark set is encoded, makes this a much easier task (human time is expensive, computer time is not). This is also a way to prevent human errors as much as possible.

A similar reasoning applies to convergence testing: if it can be automated via workflows, then more people will do it, and will do it properly. Such a(n automated) procedure to determine numerical settings is to be preferred over providing default settings only, as it can take specific details of an individual case into account.

Having 'all' published DFT results (and also non-published ones) properly stored in databases, not only their results but also the input and output files, would be a rich source of information. It helps carrying over expertise to other users, it helps detecting problems, mistakes or fraud in published data, it allows reusing, recycling or re-analyzing existing data,... It just speeds up research a lot.

Disadvantages of uploading your data to a database are: Which database should I use? Will it survive for sufficiently long? Does it accept the format in which my data are? How easily can it be searched? The Optimade project ([www.optimade.org](http://www.optimade.org)) aims to make all databases interoperable, which alleviates to some extent the pain of having to make a choice.

Possible dangers are that contextless data can be dangerous for non-experts. For instance, a PBE band gap can easily be interpreted by a non-expert as a 'correct' band gap. Or a fast benchmark procedure with e.g. given a coarse k-mesh can only be compared with data obtained using the same coarse k-mesh. Information and warnings can be added, but this never prevents all incorrect use. In addition, too much information may have a deterrent effect.

Finding conflicting information in different databases (or even in the same database) can feel as disturbing. On the other hand, it's better to know there is this spread, rather than using a single value and believing out of ignorance it is the correct one.

---

*This is a digest of a dedicated discussion session held at PQ-DFT 2019. For other digests, videos of all talks and summarizing recommendations, please visit <https://pqdft2019.abinit.org/>. To access the videos directly on Youtube, visit <http://bit.ly/2XFKUCL>. Any comments, thoughts or items you want to discuss? Feel free to contact Stefaan Cottenier ([stefaan.cottenier@ugent.be](mailto:stefaan.cottenier@ugent.be)) or Kurt Lejaeghere ([kurt.lejaeghere@ugent.be](mailto:kurt.lejaeghere@ugent.be)).*