

3.6: Special Notes on Using Public Chemical Databases

Availability and Data Exchange

All the databases mentioned in section 3.4 and 3.5 (including PubChem) are public databases that provide their contents free of charge, and in many cases they also provide a way to download data in bulk and integrate them into one's own database. Therefore, it is very common that database groups exchange their information with each other. This often raises some technical concerns. For example, different databases may use different chemical representations to refer to the same molecule. This may result in incorrect chemical structure matching between the databases, leading to incorrect data integration. In addition, when one database has incorrect information, this error often propagates into other databases. The error propagation issue is a serious, but very common, problem.^{1,2} Therefore, when using information in these databases, one should keep in mind various data accuracy and quality issues prevalent in these databases. A goal of this course is to help students develop the ability to critically assess chemical information available in public databases.

References

1. Schnoes, A. M.; Brown, S. D.; Dodevski, I.; Babbitt, P. C. *PLoS Comput. Biol.* **2009**, *5*, e1000605.
2. Philippi, S.; Kohler, J. *Nat. Rev. Genet.* **2006**, *7*, 482.

3.6: Special Notes on Using Public Chemical Databases is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.