Recent activities around the crystallography open databases COD, PCOD and P2D2

Armel Le Bail
Université du Maine, Laboratoire des Oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France, Email : armel.le_bail@univ-lemans.fr

Keywords : open databases, crystal structures, prediction, powder diffraction

The Crystallography Open Database (COD) [1] is a very recent online searchable tool (www.crystallography.net, founded February 2003) offered to the scientific community. Five years after its foundation, the COD passed a major milestone in 2008, by archiving the 50 000th entry (containing ~10 000 entries from the AMCSD [2], and including everything but large molecules like proteins or nucleic acids). This number is estimated to be less than 10% of the total expected which is currently increasing by ~40 000 crystal structures published annually. To attain completion some day, the COD should thus add much more than 40 000 new entries per year, which is far from being realized yet. In December 2003, a subset of the COD was created and named PCOD (Predicted Crystallography Open Database) in an attempt to gather all these theoretical crystal structures which will very probably soon exceed in number the really observed ones. The PCOD contains more than 100 000 virtual phases (N- and N-N'-connected 3D nets) predicted with the help of the GRINSP software [3]. In February 2007, from the PCOD data, was created the P2D2 (Predicted Powder Diffraction Database) [4] allowing for identification by search-match procedures in a way similar as done from the PDF (Powder Diffraction File) [5]. In September 2007, at its recent meeting, the IUCr Executive Committee decided that the CIFs associated with structural papers published in IUCr journals should be made freely available to the databases, including COD. This welcomed decision clearly allows for a new departure of the COD of which the center of decisions has been transferred from Le Mans (France) to Vilnius (Lithuania) in December 2007. We can consider that the COD petition in favour of crystallography open databases has attained some of its targets, after the accumulation of more than 1800 signatures. It remains to obtain a permission to copy the CIFs available as supplementary materials at the ACS and RSC journal articles web sites, and to convince more crystallographers to deposit their crystal structures, real or virtual as soon as they are determined or simulated, published or not. Beside the scheduled increase of the number of entries, new developments are now undertaken by the COD team, mainly at Vilnius or elsewhere, which will be described, including automatic process for data deposition, validation and correction, new search interface, mirroring and synchronization, powder pattern computation. The uses of COD already realized in external database interfaces will also be listed, including the integration of the COD data inside of the Crystal Impact Match software [6] for identification through calculated powder patterns, the development of an user interface [7] to the COD and PCOD allowing for visualization, importing, exporting data to other applications. These realizations are the clear consequence of the open character of the COD data. The COD Advisory Board (see ref. [1] authors, adding S. Gražulis, Y. Matsushita, P. Moeck, M. Quirós Olozábal) encourages the scientific community to extend their uses of these open tools in an unlimited way.

[1] Le Bail A., Chateigner D., Chen X.L., Ciriotti M., Cranswick L.M.D., Downs R.T., Lutterotti L., Yokochi A.F.T., Acta Cryst. **A61**, C481 (2005).
[2] Downs R.T. and Hall-Wallace M., American Mineralogist **88**, 247-50 (2003).
[3] Le Bail A., J. Appl. Cryst. **38**, 389-395 (2005).
[4] Le Bail A., Powder Diffraction, in the press (2008).
[5] Kabekkodu S.N., Faber J. and Fawcett T., Acta Cryst. **B58**, 333-337 (2002).
[6] Match! – Crystal Impact – http://www.crystalimpact.com/match/match17.htm
[7] Fennick J.R., Brandon Keith J., Leonard R.H., Truong T.N. and Lewis J.P., J. Appl. Cryst. **41**, 471-475 (2008).