

June 30, 2003

Crystallography Open Database
<http://www.crystallography.net/>
Email : cod@cristal.org

Dear IUCr Executive Committee Members,

In August 2003, the Advisory Board of the Crystallography Open Database (COD) will request permission for unlimited access to the IUCr CIFs. The CIFs are presently available by free access at the IUCr journals web server. With this letter, the COD Advisory Board Members wish to explain their motivations, and why and how they believe that this new global database will become an invaluable tool for the International Community of Crystallographers.

Our IUCr President writes in the Newsletter 10-4 (2002): "*Emerging nations can benefit from the use of the powerful techniques of X-ray crystallography in order to analyze, understand and use the unique natural resources within their countries whether mineralogical, chemical, or biological in nature*". As stated in the first letter sent by the COD Advisory Board Members to Bill Duax (April 2003), this generous wish could more easily be achieved through free access to a global crystallography open database (including all inorganic, metallic, organometallic and organic compounds) which is the purpose of our new COD web server located at <http://www.crystallography.net/> offering atomic coordinates in the IUCr CIF files format, together with a search engine.

The initial position of COD was to develop the database through individual or laboratory-based contributions. Rather than systematically copying the atomic coordinates from the literature, individual authors and laboratories are encouraged to send their own CIFs. Thanks to a few major donators (the American Mineralogist Crystal Structure Database for >4000 entries, the Institut de Physique de la Matière Condensée -Grenoble- for >1200 entries, the CRISMAT -Caen- for >850 entries and the Laboratoire des Fluorures -Le Mans- for >450 entries), as well as to many individual contributors, the COD already contains more than 12.000 entries. This is, however, only 3% of the 400.000 known published structures. Permission from IUCr would more than double the current COD content.

Currently, a search in COD is completed by two powerful options:

1. Combining in any logical way: text, cell volume, chemical elements and strict number of elements.
2. Cell parameters ranges a_{\min} - a_{\max} , b_{\min} - b_{\max} , etc.

We believe that one of the first needs of a crystallographer trying to solve and publish a new crystal structure is to be able to verify that he is not working on a structure that has already been determined. This is easily accomplished by a check of cell parameters and approximate chemical formula. This is also one of the first needs of a reviewer in charge of the evaluation of a manuscript. These above search possibilities should fulfil these basic needs.

We will ask the IUCr for one of two options:

1. Permission to include CIF data in the COD database.
2. Permission to construct REF files from the IUCr's CIFs that include the reference, chemical formula, cell parameters and space group. Then a search on COD that returns a structure published in IUCr journals will return a hyperlink to the CIF at the IUCr web server. Thus the CIF does not need to be on the COD server side since it is already freely available at the IUCr

journals server side. In this way, COD can serve as a search engine to the IUCr journals. As a test, the COD is already pointing at the IUCr web site for the Acta Cryst. C 1991 and 1992 CIFs.

Using COD, a crystallographer (or a reviewer) working in interdisciplinary fields, in both the organic and inorganic worlds, would find fast answers to questions at one unique entry point (the COD server) rather than having the obligation to buy all the fragmented parts of known crystal structure databases (e.g. CSD for organics and organometallics, ICSD for inorganics, CRYSTMET for metals and intermetallic compounds) and rather than to have to search successively in all three or two of them if the chemical formula is unknown or dubious. The proteins and nucleic acids databases, PDB and NDB, are already free access, and their inclusion into COD will not be considered. There are few chances to make any protein or nucleic acid crystallized compound by pure haphazard, whereas an inorganic compound can well occur when expecting an organometallic one.

There are some data quality issues that need to be considered. Currently, COD expects a donor to upload to the COD only his best data (CIFs as built by the refinement software, and thus free of typos at least on atomic coordinates and cell parameters). We assume that the donor of the data does the quality check. The COD is confident that Crystallographers will demonstrate a high degree of professionalism and autonomy.

About durability of the COD, the current ten Advisory Board Members offer some guarantee. The open nature of the COD, which can be downloaded either in part or completely, including all data and source code, is also an argument favouring its survival. The whole COD system is based on open source software (Apache/MySQL/PHP) and can be used by interested parties (contact cod@crystal.org). It is simple to install and would be quite useful to laboratories wanting to build a crystal database from their own crystal structure determinations, for either an intranet or an internet external access (or for building any crystal subset of interest to specialized Crystallographers such as zeolites, phosphates, fluorides, nitrates, carbonates, sulphates, etc). Collecting all such laboratory fragments in the global COD would offer an invaluable service to the world crystallographic community and particularly in emerging nations. The CRYSTMET, ICSD and CSD databases offer much more complete and powerful services than COD, which is limited to the minimum information allowing a crystallographer to survive at low cost. Even crystallography journals (not only IUCr journals) may decide in the near future to open their own database of published crystal structures in a searchable way. The journals interest is to sell articles, so that it seems clear that they may decide to open a database limited to REF files, the client wanting atomic coordinates would have to buy a copy or the published paper. REFs may include compounds in the COD that were not completely structurally characterized (no atomic coordinates, but with chemical formula, cell parameters and space group proposed by isotypy with a fully characterized analogue, etc). In this way, the COD would even be more complete than databases distributed for fees.

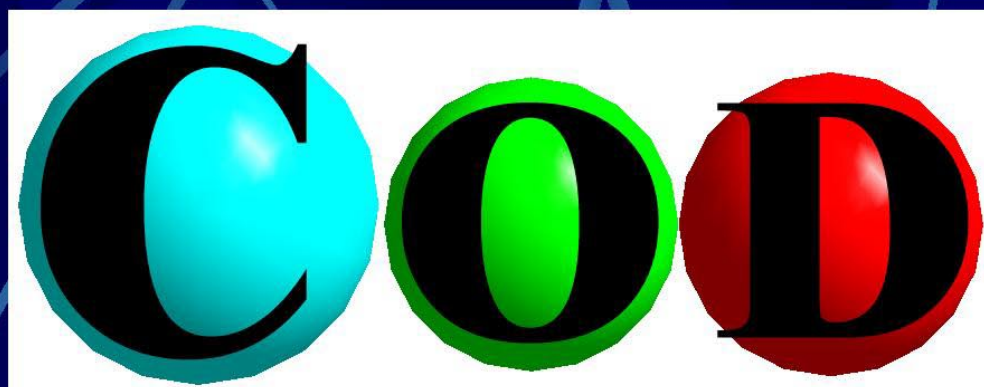
With this letter, we have presented to you a vision of free and easy access to the crystallographic data via the Internet. Once the database has developed to some maturity, it is our dream that its use will promote the science of crystallography at all levels of education and to all parts of the world. We envision the use of software, intimately associated with CIFs that will promote and ease the interface to understanding our exciting profession. We hope that as members of the IUCr executive committee, you can share our vision and grant COD access to the data files published in the IUCr journals. For more details about the COD, REF files and so on, visit the web site: <http://www.crystallography.net/>

Sincerely yours,

The COD Advisory Board :

M. Berndt (Germany), D. Chateigner (France), X.L. Chen (China), M. Ciriotti (Italy), L.M.D. Cranswick (Canada), R.T. Downs (USA), A. Le Bail (France), L. Lutterotti (Italy), H. Rajan (USA), A.F.T. Yokochi (USA).

Crystallographers join the Crystallography Open Database



www.crystallography.net



Deposit your crystal data
in the Public Domain
Thanks !



Advisory Board :

Michael Berndt, Daniel Chateigner, Xiaolong Chen, Marco Ciriotti, Lachlan M.D. Cranswick,
Robert T. Downs, Armel Le Bail, Luca Lutterotti, Hareesh Rajan, Alexandre F.T. Yokochi