Preface

F. H. Allen[a] and J. P. Glusker[b]

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, and [b]The Institute for Cancer Research, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

Fifteen years ago the first monograph on crystallographic databases [Allen, F. H., Bergerhoff, G. & Sievers, R. (1987). *Crystallographic Databases*. Chester, England: International Union of Crystallography] was prepared for the 1987 IUCr Congress in Perth, Australia. A total of 95 684 crystal structures were recorded at that time in the various structural databases that store three-dimensional atomic coordinates, as listed in Table 1. The intervening period has seen a dramatic upsurge in the number of reported single-crystal analyses and a new database, the Nucleic Acid Database, was introduced in 1991. Current statistics (Table 1) now show an overall fourfold increase (to a total of 407 979) in the number of structures in the public domain. Clearly, the most dramatic figure in Table 1 is the 44-fold increase in the holdings of the Protein Data Bank, a quantification of the enormous strides that have taken place in macromolecular crystallography in the last two decades, evidence of the encouragement of data deposition and a demonstration of the vital importance of three-dimensional protein structures to modern science.

Table 1 also shows that the structural databases cover the complete spectrum of the types of chemical compounds, and the statistics show that the crystallographic technique has truly moved from the method of last resort to the method of choice for structural characterization. The value of crystallographic information is not, however, restricted to structure determinations from single crystals. The recording and storage of powder diffraction patterns has a long and distinguished history, and forms the basis for the rapid identification of materials generated in a wide variety of industrial and academic environments. Structure analysis from powder diffraction data is now feasible.

Two key areas of database usage have shown very significant progress, fuelled by the enormous recent advances in computing and communications technology. First, the acquisition, processing, validation and dissemination of data has been revolutionized by technology and by the advent of the Crystallographic Information File format (in its various guises), which now provides the essential international standard for the electronic interchange of data between authors, journals and the databases. Secondly, there have been continuous developments of software systems for the interrogation and scientific applications of data retrieved from all of the databases. These software developments have themselves had two effects: first, they have significantly simplified database access and accelerated the more widespread use of the databases in the broader scientific community; secondly, the requirements of an increasing user-base has fed back into further software developments.

This special issue of *Acta Crystallographica* is designed to provide the reader with an overview of the current status of the major crystallographic databases, together with one or two other papers that help complete this snapshot in time. In planning the issue within the available space, an initial page allocation was agreed with each database group. This allocation was to be used to cover

(i) a basic database description,

(ii) currently (or soon to be) available software systems for database access, and

(iii) one or more papers which describe typical database applications, or which survey a broader range of scientific applications already published in the open literature.

As the manuscripts arrived it was possible to adjust these initial allocations to better accommodate the needs of each group, or to suggest a more appropriate ordering of material that fitted objectives (i)–(iii) more closely.

The papers themselves clearly fall within the scope of either Section B or Section D of *Acta Crystallographica* and have been assigned page numbers on that basis. However, all subscribers to either of these sections will receive the full composite issue, which will also be available on a stand-alone basis. The ordering of the Section B papers follows accepted norms, moving from metals and inorganics through to molecular species. Papers relating to the NDB and the PDB naturally fall within the scope of Section D, together with one paper, describing applications of the Cambridge Structural Database in the life

Table 1
Comparative entry statistics for the structural databases in 1987 and 2002.

N(ent) represents the number of entries.

Database	Earliest reference	*N*(ent) 1987 (% of total)	*N*(ent) 2002 (% of total)	Increase factor
CRYSTMET Metals, alloys	1913	11 000 (11.5%)	69 054 (16.9%)	6.27
ICSD Inorganics, minerals	1915	26 045 (27.2%)	59 285 (14.5%)	2.28
CSD Organics, metal-organics	1923	58 239 (60.9%)	260 535 (63.9%)	4.47
PDB Proteins	1963	400 (0.4%)	17 637 (4.32%)	44.09
NDB Nucleic acids	1969	–	1468 (0.4%)	–
Totals		95 684	407 979	4.26

sciences, where editorial judgement indicates Section D to be the most appropriate assignment.

We hope that this issue will prove valuable to crystallographers and non-crystallographers alike in providing a comprehensive survey of the activities of the various database groups and, particularly, in illustrating the wide-ranging information retrieval and research possibilities provided by the crystallographic databases. We take this opportunity to thank all of those who have contributed to the issue, and also those scientists who, often at very short notice, were kind enough to provide informed and valuable opinions on the submitted manuscripts.