Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design

Alec Belsky,[a] Mariette Hellenbrandt,[b] Vicky Lynn Karen[a]* and Peter Luksch[b]

[a]National Institute of Standards and Technology, Gaithersburg, Maryland 208999-8520, USA, and [b]Fachinformationszentrum Karlsruhe, D-76344 Eggenstein-Leopoldshafen, Germany

Correspondence e-mail: vicky.karen@nist.gov

Received 15 April 2002
Accepted 17 April 2002

The materials community in both science and industry use crystallographic data models on a daily basis to visualize, explain and predict the behavior of chemicals and materials. Access to reliable information on the structure of crystalline materials helps researchers concentrate experimental work in directions that optimize the discovery process. The Inorganic Crystal Structure Database (ICSD) is a comprehensive collection of more than 60 000 crystal structure entries for inorganic materials and is produced cooperatively by Fachinformationszentrum Karlsruhe (FIZ), Germany, and the US National Institute of Standards and Technology (NIST). The ICSD is disseminated in computerized formats with scientific software tools to exploit the content of the database. Features of a new Windows-based graphical user interface for the ICSD are outlined, together with directions for future development in support of materials research and design.

1. Introduction

Access to crystal structure data can be a key step in solving research and applications problems involving materials, as in the chemical (catalytic materials), petroleum (zeolites) and electronics (epitaxic growth and thin films) industries. These data are of interest to analysts in areas such as materials design, property prediction and compound identification. Access to reliable information on the structure of crystalline materials helps researchers concentrate experimental work in directions that optimize the discovery process.

As in other disciplines, the field of chemistry has undergone significant changes in recent years. New instrumentation, better electronics and computers, and a wide variety of synthetic methods have all increased our ability to observe, control and manipulate materials on an atomic scale. With new synthetic techniques, such as nanostructure fabrication, sol–gel methods and molecular templates, the chemist is no longer limited to growing only nature's preferred materials, but can now assemble metastable phases in the hope that one may have unusual physical properties.

A second approach uses automated combinatorial syntheses – one can now make hundreds of compounds at a time and *then* discover what properties the materials do or do not have. Although this combinatorial approach has traditionally been associated with drug discovery, the method is increasingly being directed towards the formation of new catalysts and electronic materials. The combination of advanced physical theory, applied mathematics and high-speed computing form the basis for the modeling and simulation

Table 1
Number of crystal structures archived annually to the ICSD since 1992.

Year	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Entries	1737	1949	2122	2326	2776	2881	3182	4277	6109	5589

algorithms that are beginning to impact materials design. Active areas include the use of *ab initio* (Hartree Fock) calculations for predicting crystal structures and for simulating phase transitions, while advanced mathematical techniques such as finite element analysis are being used to model multiparticle crystalline systems.

A third major factor influencing the way we do chemistry today has been the development of databases and the emergence of chemical informatics. A relatively new development within the past 25 years, scientific databases are changing the way that chemistry is being done – research can be carried out on individual properties or collections of properties.

Designing the macroscopic properties of a material by selecting and modifying variously scaled building blocks is a relatively new approach in chemistry. We can think about materials and their properties from the lowest levels of atoms, clusters and aggregates, and extend these up to the level of bulk materials. The databases required to support the various levels are different and, consequently, the data products built in the future will be considerably different. Ideally, models on different levels must work together, although major conceptual and theoretical work still needs to be done in order to close the gap between the various scales.

Two approaches are being used in materials research and design. Both assume that there is a correlation between the properties of a material and its chemical or crystallographic structure. The first approach uses experimental data to identify the effects of a structural change on a given property and then expresses the relationship using a mathematical model. The second approach *starts* with a model or theory and attempts to design a structure *ab initio* so that the material will perform in a specific way. With either approach, the goal is to construct, and narrow down, the list of candidate materials to be synthesized and tested. Informatics tools such as knowledge-based software, random search methods and computational neural networks have all been used to help construct lists of candidate materials. Predicting properties prior to their synthesis remains the critical new task of computer-assisted design.

What are the demands that the materials community is placing on crystallographic data activities? The earliest uses of crystallographic databases typically focused on one entry at a time, for purposes of identifying unknown phases and analyzing isostructural materials. Gradually, the demand began to shift towards using the entire database in order to speculate, predict and correlate structures with observed functions. Here, examples are plentiful, *e.g.* through links between:

(i) unit cells and crystal systems,

(ii) space-group symmetry and optical or magnetic properties, and

(iii) crystal packing and catalysis or molecular separation.

Now that we have moved into the simulation and modeling phase, the driving forces are just being felt. Crystallographers and materials designers alike are beginning to demand and develop a new layer of software for the intelligent access of scientific data. Lattice theory and similarity searching, algorithms to fine-tune valences and chemical composition, and tools to create structure–property and structure–stability maps should all have immediate application when manipulating structures and matching interfaces between components. Clearly, crystallographic databases have a fundamental role and there is an increasing demand for data which is more accurate, more comprehensive and more accessible. All of these points are illustrated in a growing list of publications that use the ICSD as the fundamental data source for applications research (see *e.g.* Braga *et al.*, 2000; Garcia-Rodriguez *et al.*, 2000; Bergerhoff *et al.*, 1999; Smrcok & Mailing, 1999; Sokolowski, 1999; Malinovskii, 1998; Sokolowski & Hodorowicz, 1998; Nyfeler & Armbruster, 1998; Vegas & Isea, 1998; Wolverton & Zunger, 1998; Ceder, 1998).

2. The Inorganic Crystal Structure Database (ICSD)[1]

2.1. History

For almost two decades, Fachinformationszentrum Karlsruhe (FIZ) has been cooperatively producing the ICSD – first with the University of Bonn, then with the Gmelin Institute from 1990 to 1998 and, since 1997, with the US National Institute for Standards and Technology (NIST). The ICSD is a comprehensive collection of crystal structure information for non-organic compounds including inorganics, ceramics, minerals and metals. The ICSD covers the literature from 1915 to the present. Recent efforts have been focused on modernizing and evaluating the ICSD. This work has included a complete redesign of the ICSD database structure, conversion and loading of the data into a relational database management system, designing graphical user interfaces to access the data and creating scientific application modules to analyze the results of a database search.

2.2. Information content

The ICSD currently contains more than 60 000 entries. In order to be included in the database, the structure has to be fully characterized, the atomic coordinates determined and the composition fully specified. A typical entry includes, *inter alia*, the chemical name, formula, unit cell, space group, complete atomic parameters (including atomic displacement parameters), site occupation factors, title, authors and literature citation. In addition to the published data, many items are added through expert evaluation or are generated by

[1] See also: http://www.fiz-informationsdienste.de/DB/icsd/index.html.

computer programs, such as the Wyckoff sequence, molecular formula and weight, ANX formula, mineral group *etc.*

The number of entries archived annually to the ICSD has been steadily increasing over the years, as shown in Table 1. Data are either keyboarded from the original publication or converted from CIF files (Hall *et al.*, 1991; http://www.iucr.org/iucr-top/cif/spec/) provided by the publisher or other cooperating organizations. A total of 1166 journal titles are covered in the ICSD and Fig. 1 shows the 12 most productive journal titles and the number of articles from each that are currently available in the database.

2.3. Data evaluation

Prior to their inclusion in the ICSD, the data are examined for completeness and various consistency checks are performed. A detailed description of the various checks is given by Bergerhoff & Brown (1987), Behrens (1996), Bergerhoff *et al.* (1996) and Fluck (1996). As entries are added to the ICSD, the data are evaluated by experts in specific disciplines and by specialized computer programs. Several types of evaluation are performed, including examination of individual data items and looking for consistency within a complete entry. The crystallographic evaluation to determine the relationship of an individual entry to the entire database continues routinely. Scientific strategies and code are used to identify which entries may represent related or duplicate crystal structure determinations. Database searches have been carried out to locate entries which are identical with respect to reduced cell parameters, space groups, Wyckoff positions and molecular formulae, among other criteria. Various subsets of the data were prepared and examined in detail by crystallographic experts. This expert evaluation is an on-going data activity.

3. The ICSD in transition: migration to a relational database system

The Inorganic Crystal Structure Database has been completely restructured, as illustrated in Fig. 2. The existing column-oriented ICSD formats have been converted into a relational database design which includes more than 200 fields and 25 tables to describe a structure. Use of a modern database management system is critical as it provides a flexible and expandable foundation for both stand-alone and online data products.

During this transition, improvements have been made to the content and accessibility of data items. For example, mineral names and associated group names are now both included in the ICSD; journal names and codens will be associated with the International Standard Serial Number (ISSN) more commonly used by publishers; and space groups and symmetry record tables have been edited and will be redesigned. While maintaining CIF-compatibility, additional functionality is being added through the calculation of derived data items, such as reduced and standard cells. In addition, more than 8000 entries have been re-evaluated on an individual basis to establish the relationships of a single entry to the entire database, the set of minerals entries has been evaluated, and numerous corrections have been made throughout using both scientific and database management system validation criteria.

Although it was a considerable effort to write conversion and data generation routines in order to import the data into relational tables, the new system should improve ICSD responsiveness to the user community by allowing greater flexibility in data maintenance and validation as well as portability to other systems, as indicated in Fig. 2.

4. Graphical user interface

A major effort has been made to create a Windows-based graphical user interface for the ICSD. This software product is tabular in design (Figs. 3*a*–*h*) and allows for searching in five general categories: Chemistry, Crystal Data, Reduced Cell, Symmetry and Reference Data. The new software includes enhanced features for the characterization of materials based on lattice search and chemistry search modules, and provides three-dimensional visualization and powder pattern simulation for inorganic structures.

In the Chemistry search, we have designed what we refer to as an 'exclusive OR' (Fig. 3*a*). When a first attempt to characterize a material fails, a researcher can usually define the components that went into the synthesis. This type of search allows the user to input several starting materials and auto-

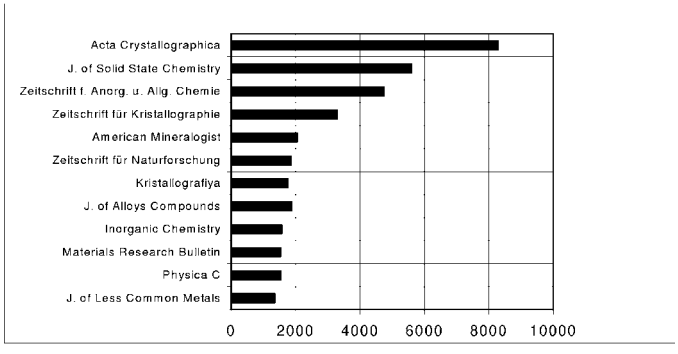

Figure 1
The 12 most productive journals contributing to the ICSD.

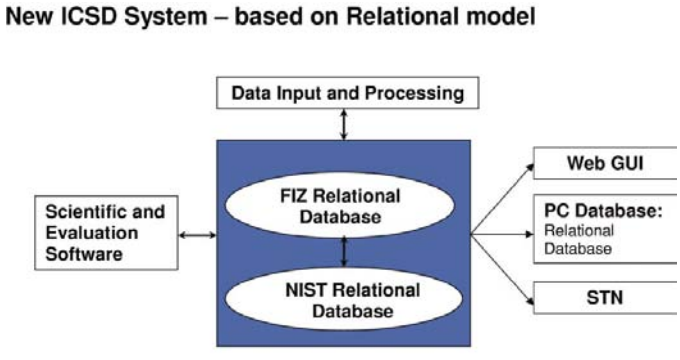

Figure 2
Transition of the ICSD to a relational model.

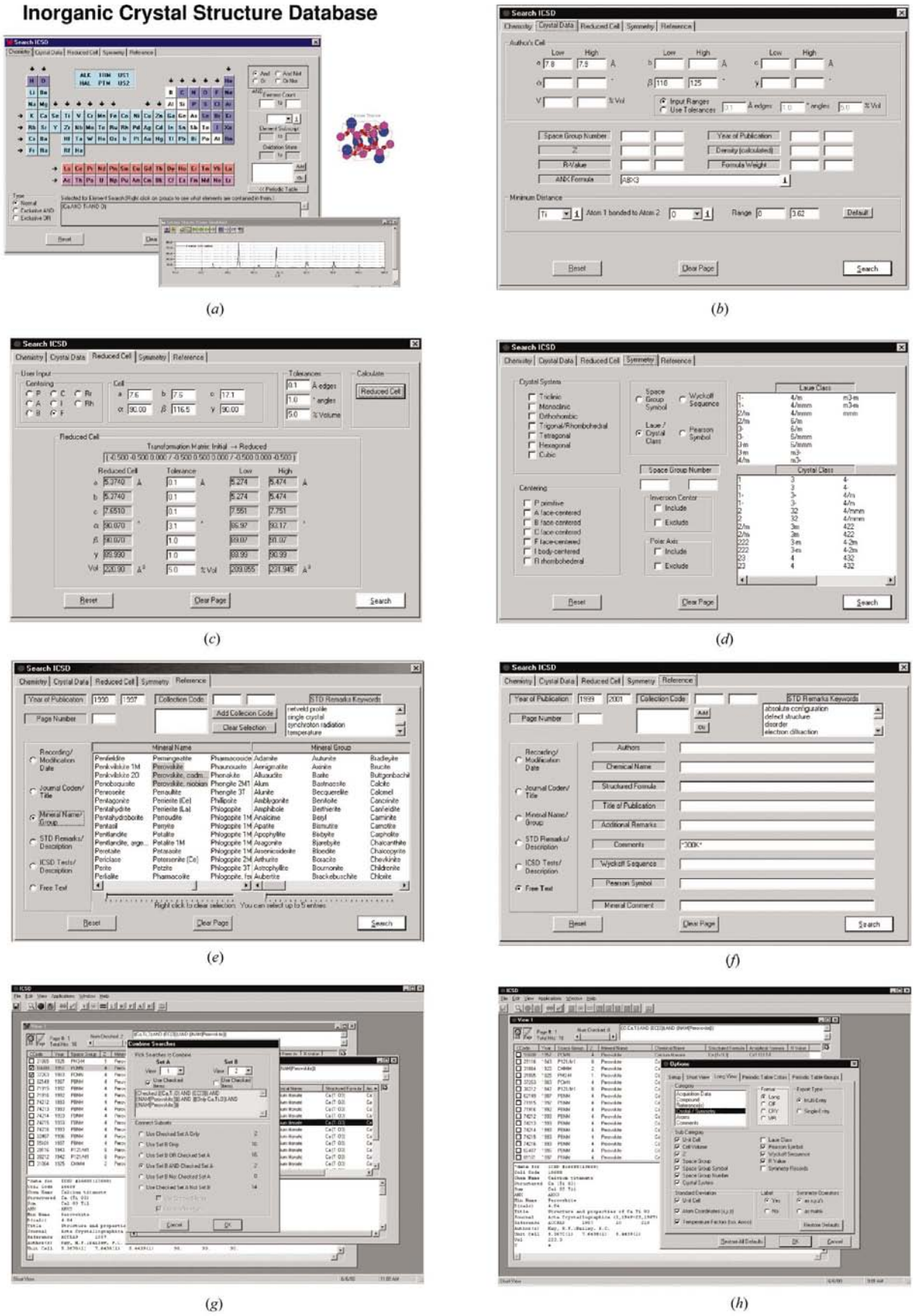

Figure 3
The new PC Windows graphical user interface for the ICSD based on the relational database structure. Individual screens are described in the text.

matically search on all possible combinations of the elements rather than trying to input and search each combination by hand. Once a search has been made, the user can generate a three-dimensional view of the structure (Fig. 3*a*, right), calculate powder patterns (Fig. 3*a*, bottom frame) or export the information into other applications.

The Crystal Data screen (Fig. 3*b*) allows searches such as structure type and author's cell. A lattice-based search (Fig. 3*c*) has been added to allow the user to input an experimentally determined cell, calculate the reduced cell (by clicking the button in the upper right of Fig. 3*c*) and then search the ICSD. This is possible because the archival database has been augmented to include the standard reduced cell. All symmetry properties are searchable (Fig. 3*d*), such as space group, Wyckoff position and Pearson symbol. In the Reference screen (Fig. 3*e*) the mineral name, mineral group and various keywords such as temperature, pressure or type of experiment can be searched; all text fields are searchable (Fig. 3*f*), including the title of the publication, authors names and additional remarks.

The ICSD can be searched in a hierarchical mode or in Boolean mode using AND, OR, NOT logic and multiple result sets can be generated. The different result sets can be combined (Fig. 3*g*) and individual entries selected by the user to customize the results for a specific problem. Finally, there is extensive flexibility in exporting the data into user-defined and standard formats (Fig. 3*h*). These options help the user export the data to outside applications.

5. Availability

Currently the database is available under different retrieval systems and platforms. The new Windows version with additional search functionality and new visualization facilities for three-dimensional structure representation has been described above.

The ICSD has a web-accessible interface (Hewat, 2002) developed by a cooperation between FIZ–Karlsruhe and the Institut Laue–Langevin (ILL), Grenoble, France. The database can be searched *via* a simple graphical interface or in the expert mode which follows exactly the CRYSTIN command syntax. The command-oriented Fortran package can still be installed under UNIX operating systems, but will not be supported in future. In addition, the ICSD is also accessible online *via* STN International using the 'STN on the WEB' interface or 'STN Classic'. The ICSD contains the *CDIST* software package to calculate and display interatomic distances and angles. Bibliographic information, property information and indexing terms are all searchable.

The ICSD is directly licensed to private industry, universities and individual scientists; in addition, there are numerous indirect applications through instrument manufacturers, software vendors and third party distributors. Interoperability of the ICSD with other data sources is becoming increasingly important for future development.

6. Future directions

Modern data structures and interfaces can support the creation of links or cross-references between different sources of data. The system design and underlying data structures used for the ICSD can accommodate different databases, each with different scientific, application, business and security requirements; multiple window navigational features allow the integration of domain specific analysis and predictive tools to be developed. First attempts have been made to link ICSD entries to the primary journal articles in which they have been published. The link to the crystallographic journals published by the IUCr is already working in the Web version of the ICSD. This new feature will allow ICSD users to access the electronic version of the original article, provided that they have subscribed to the electronic journal. The link is intended to be bi-directional, *i.e.* the reader of the electronic journals

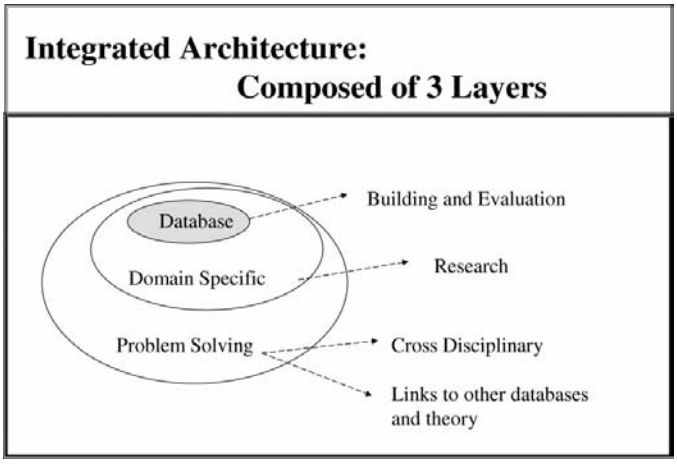

Figure 4
Components of a distributed database system for materials research and design.

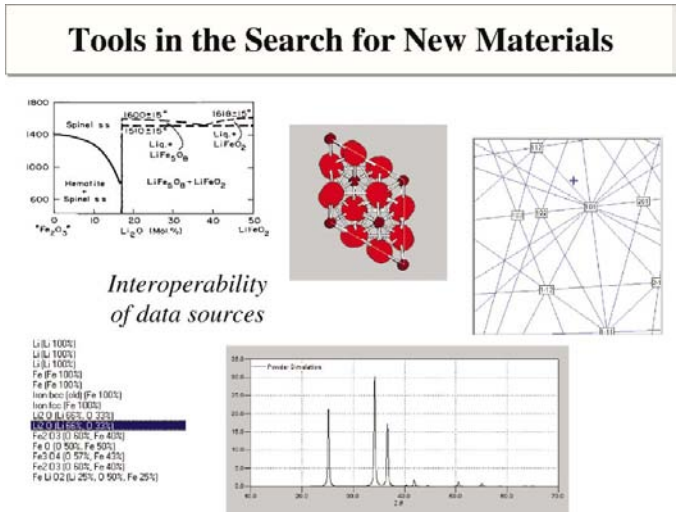

Figure 5
A central role for the ICSD in the search for new materials.

will obtain the accession number of the corresponding structure in the ICSD.

The database system and software design for the ICSD database is, in principle, the same for all materials databases (Fig. 4). At the core is a distributed database containing both archival and derived data items. The ICSD graphical user interface represents the intermediate, domain-specific layer – here is where the research and scientific expertise is important. Finally, there is an outer problem-solving layer that cuts across disciplines and links various property databases and theory.

Interoperability of data sources is critical because it will allow the user to start in one discipline or experiment – whether single-crystal diffraction or phase diagrams – and retrieve all the information available in other materials databases (Fig. 5). For example, the ICSD has been incorporated into an electron diffraction instrument for phase identification. Once a sample has been analyzed by electron diffraction, giving an elemental composition and a backscatter electron diffraction pattern, the database is searched by element type and a simulated pattern is calculated, such as the one with the indexing numbers on the far right of Fig. 5. A phase can be identified automatically by matching the calculated and observed patterns. Selective cross-linking of the crystallographic data with other database systems is an important step in realising the potential envisioned by today's materials designers. This interoperability with other data sources and software tools appears to be one of the emerging driving forces for innovation today.

The authors wish to acknowledge substantial efforts by its partners and others working on the production and evaluation of the ICSD including: Dr Heinrich Behrens, formerly of FIZ–Karlsruhe, and Shari Young of the NIST Standard Reference Data Program. Special thanks are due to Professor Dr R. Allman for his work in checking and correcting a large number of ICSD entries and for his evaluation of mineral names.

References

Behrens, H. (1996). *J. Res. Natl Inst. Sci. Technol.* **101**, 365–373.
Bergerhoff, G., Berndt, M. & Brandenburg, K. (1996). *J. Res. Natl Inst. Sci. Technol.* **101**, 221–225.
Bergerhoff, G., Berndt, M., Brandenburg, K. & Degen, T. (1999). *Acta Cryst.* B**55**, 147–156.
Bergerhoff, G. & Brown, I. D. (1987). *Crystallographic Databases*, edited by F. H. Allen, G. Bergerhoff and R. Sievers, pp. 77–95. Chester: International Union of Crystallography.
Braga, D., Maini, L., Grepioni, F., Mota, F., Rovira, C. & Novoa, J. J. (2000). *Chem. Eur. J.* **6**, 4536–4551.
Ceder, G. (1998). *Science*, **280**, 1099–1100.
Fluck, E. (1996). *J. Res. Natl Inst. Sci. Technol.* **101**, 217–220.
Garcia-Rodriguez, A., Rute-Perez, A., Pinero, J. R. & Gonzalez-Silgo, C. (2000). *Acta Cryst.* B**56**, 565–569.
Hall, S. R., Allen, F. H. & Brown, I. D. (1991). *Acta Cryst.* A**47**, 655–685.
Hewat, A. (2002). http://barns.ill.Fr/dif/icsd/ and http://www.Fiz-informationsdienste.de/DB/icsd/www-recherche.html.
Malinovskii, T. I. (1998). *Cryst. Rep.* **43**, 241–255.
Nyfeler, D. & Armbruster, T. (1998). *Am. Mineral.* **83**, 119–125.
Smrcok, L. & Mailing, J. (1999). *Powder Diffr.* **14**, 5–9.
Sokolowski, J. A. (1999). *J. Alloys Compd*, **286**, 219–223.
Sokolowski, J. A. & Hodorowicz, D. (1998). *Chem. Papers (Chem. Zvesti)*, **52**, 136–139.
Vegas, A. & Isea, R. (1998). *Acta Cryst.* B**54**, 732–740.
Wolverton, C. & Zunger, A. (1998). *J. Electrochem. Soc.* **145**, 2424–2431.