New Powder Diffraction File (PDF-4) in relational database format: advantages and data-mining capabilities

Soorya N. Kabekkodu, John Faber* and Tim Fawcett

International Centre for Diffraction Data (ICDD), 12 Campus Boulevard, Newtown Square, Pennsylvania 19073-3273, USA

Correspondence e-mail: faber@icdd.com

The International Centre for Diffraction Data (ICDD) is responding to the changing needs in powder diffraction and materials analysis by developing the Powder Diffraction File (PDF) in a very flexible relational database (RDB) format. The PDF now contains 136 895 powder diffraction patterns. In this paper, an attempt is made to give an overview of the PDF-4, search/match methods and the advantages of having the PDF-4 in RDB format. Some case studies have been carried out to search for crystallization trends, properties, frequencies of space groups and prototype structures. These studies give a good understanding of the basic structural aspects of classes of compounds present in the database. The present paper also reports data-mining techniques and demonstrates the power of a relational database over the traditional (flat-file) database structures.

1. Introduction: growth of the Powder Diffraction File over the past 60 years

The PDF is maintained and continually updated by the ICDD, a non-profit scientific organization dedicated to collecting, editing, publishing and distributing powder diffraction data for the identification of crystalline materials. Put simply, the PDF is a collection of single-phase X-ray powder diffraction patterns. This collection also includes calculated powder diffraction patterns based on reported crystal structures. Detailed discussions of the ICDD database structure can be found in a companion paper (Faber & Fawcett, 2002). In addition to interplanar spacings (d) and relative intensities (I), other useful data, such as synthesis, physical properties and crystallographic data, are stored in the database. Even though PDF-2, which is a flat-file database (organized as a sequential series of multiple records per entry), is compatible with efficient retrieval and search/match methods, the scientific community has requested a database that would provide for a broader range of analyses, for example improved quantitative analyses, full pattern display, bibliographic cross referencing *etc.*

These goals have now been achieved by developing the PDF in relational database (RDB) format. This new RDB format is designated PDF-4. Unlike a flat-file database, RDBs store information in several inter-related tables. This very feature of the relational database gives pliable access to the database to carry out data-mining studies and enhances the pursuit of conventional materials characterization using diffraction techniques (Faber *et al.*, 2001). The PDF-4 format

Table 1
Prototype structures corresponding to the numbers shown on the abscissa in Fig. 3.

No.	Compound	No.	Compound	No.	Compound
1	Fe_3C	19	Si_2Th	37	Al_2O_3
2	CaB_6	20	$CaTiO_3$	38	$AuCu_3$
3	N_4Si_3	21	$MoSi_2$	39	$CuFeS_2$
4	S_3Sb_2	22	Cr_3Si	40	Fe_2P
5	AlLiSi	23	Mg	41	TiO_2
6	Cr_3S_4	24	Cu	42	MnP
7	CrFe	25	CdI_2	43	La_2O_3
8	Th_2Zn_{17}	26	BCr	44	AlB_2
9	Ni_3Sn	27	Mn_5Si_3	45	$MgZn_2$
10	BFe	28	$CrNaS_2$	46	FeS_2
11	MoS_2	29	P_4Th_3	47	CsCl
12	CaC_2	30	AsNi	48	Co_2Si
13	Si_3W_5	31	$InNi_2$	49	Cu_2Mg
14	$BFe_{14}Nd_2$	32	$CaCu_5$	50	ZnS
15	SiO_2	33	$Mn_{12}Th$	51	CaF_2
16	$Al_{43}Ho_6Mo_4$	34	Mn_2O_3	52	Al_2MgO_4
17	Al_2CdS_4	35	BiF_3	53	Al_4Ba
18	Al_2Cu	36	Cu_2Sb	54	NaCl

enables one to search on several criteria and fields that are not searchable in PDF-2. In addition to the improved access to some of the fields, users can also build search criteria by combining individual search conditions using Boolean operators. The availability of logical operators for combining the search condition is very useful in arriving at the desired information from the database. The Boolean search method becomes a very powerful search method especially when combined with either Hanawalt (Hanawalt et al., 1938) or Fink (Fink, 1962) methods. The following sections discuss the data-mining and search capabilities of the PDF-4/Metals and Alloys RDB.
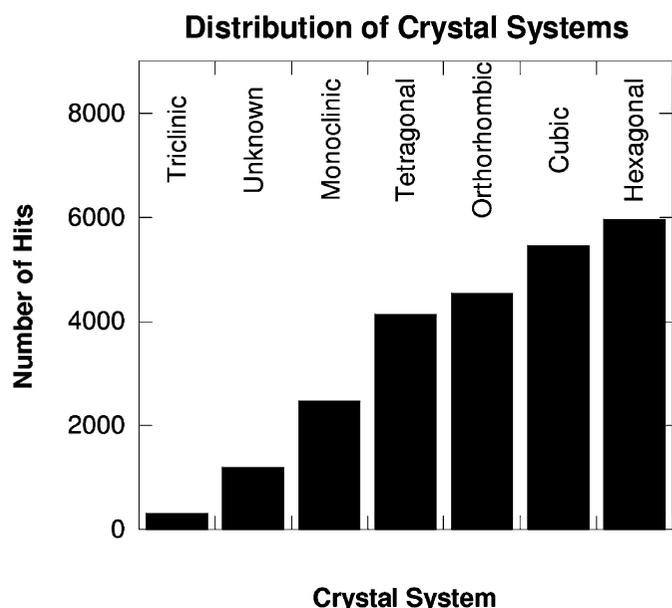
2. Symmetry trends in metals and alloys: an overview

A case study was carried out using the ICDD's PDF-4/Metals and Alloys RDB to reveal the crystallization trends in metals and alloys in terms of symmetry and structure. For this purpose, we studied the frequencies of crystal systems, space groups and prototype structures in which metals and alloys crystallize. Fig. 1 shows the number of occurrences of a particular crystal system in the PDF-4/Metals and Alloys RDB. A predominance of crystal systems with high symmetry
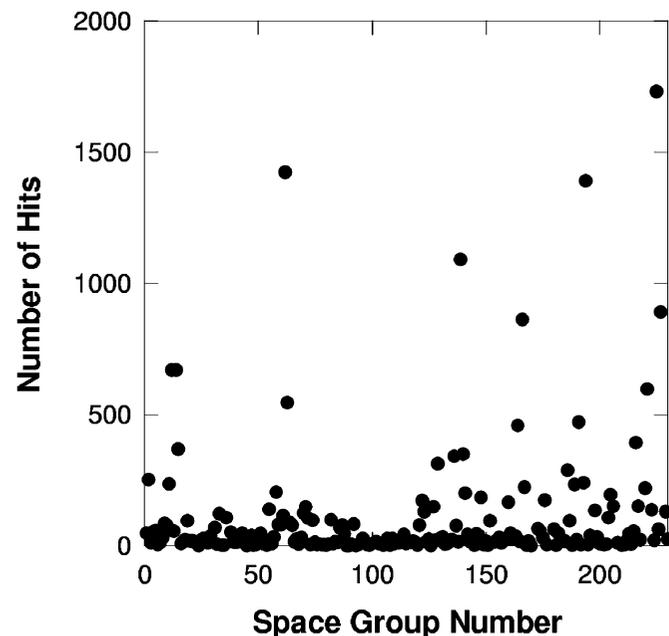
Figure 2
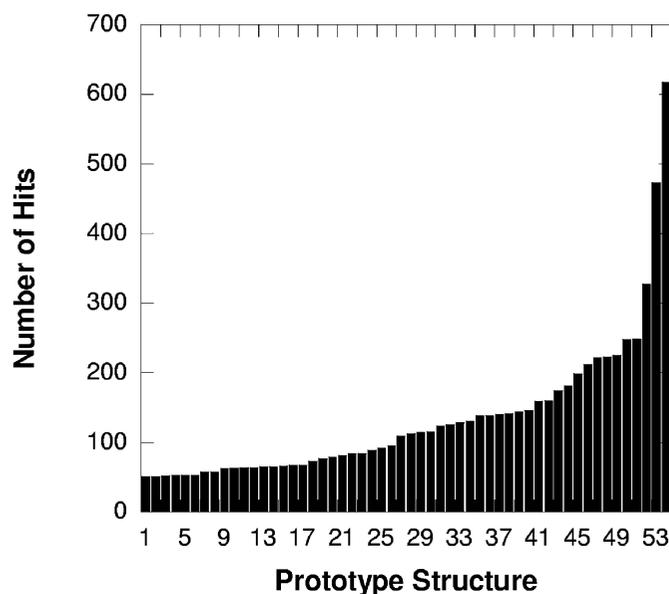Frequency of occurrence of space groups in the PDF-4/Metals and Alloys RDB.

Figure 1
Distribution of crystal systems in the PDF-4/Metals and Alloys RDB.

Figure 3
Frequency of occurrence of prototype structures in the PDF-4/Metals and Alloys RDB

Table 2
Space groups not found in PDF-4/Metals and Alloys RDB.

A space group that is centrosymmetric is marked with an asterisk.

Space group No.	Space group	Space group No.	Space group
27	$Pcc2$	171	$P6_2$
30	$Pnc2$	172	$P6_4$
37	$Ccc2$	178	$P6_122$
50*	$Pban$	179	$P6_522$
95	$P4_322$	183	$P6mm$
97	$I422$	184	$P6cc$
101	$P4_2cm$	207	$P432$
104	$P4nc$	209	$F432$
106	$P4_2bc$	219	$F\bar{4}3c$
153*	$P32_12$	222*	$Pn3n$
158*	$P3c1$	228*	$Fd3c$

is apparent. A more interesting trend is seen in terms of space groups, as shown in Fig. 2. It is clearly seen that there are a few particular space groups in which the majority of the compounds crystallize. Prototype structures give us an insight into the packing arrangement. As can be seen in Fig. 3 and Table 1, the most frequently observed prototype structures of the compounds in the database are the structure types 'NaCl' and 'Al$_4$Ba'. Fig. 3 shows the frequency of those prototype structures that occur more than 50 times. The prototype structures corresponding to the numbers on the abscissa are listed in Table 1.

Another interesting observation is that out of the 230 possible space groups, only 208 actually occur in the PDF-4/ Metals and Alloys RDB. The space groups that do not occur are listed in Table 2. As can be seen in Table 2, most of the entries that are not populated in the PDF-4/Metals and Alloys RDB are non-centrosymmetric. However, the choice of non-centrosymmetric or centrosymmetric space-group determina-
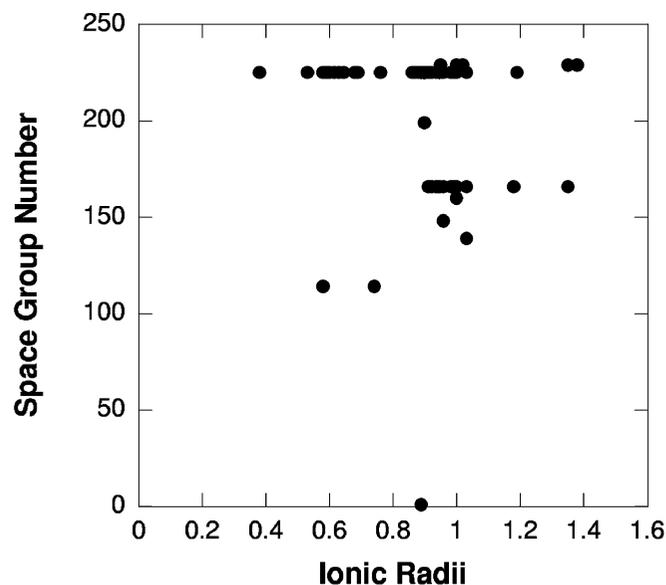
Table 3
Cell dimensions of the $R\bar{3}m$ phase of $Bi_{2-x}A_xO_{3-\delta}$ in hexagonal and rhombohedral settings.

Subscripts H and R corresponds to hexagonal and rhombohedral settings, respectively.

PDF No.	Chemical formula	a_H	c_H	a_R	α_R
410309	$Bi_{1.6}Ca_{0.4}O_{2.8}$	3.928	27.85	9.555369	23.71
480353	$Bi_{0.775}Dy_{0.225}O_{1.5}$	3.951	27.285	9.375762	24.33
480352	$Bi_{0.775}Tb_{0.225}O_{1.5}$	3.963	27.36	9.401694	24.33
480351	$Bi_{0.775}Gd_{0.225}O_{1.5}$	3.97	27.337	9.395243	24.4
410308	$Bi_{1.56}Gd_{0.44}O_3$	3.971	27.353	9.400556	24.38
480350	$Bi_{0.775}Eu_{0.225}O_{1.5}$	3.975	27.372	9.40726	24.4
410307	$Bi_{1.672}Eu_{0.328}O_3$	3.976	27.569	9.471091	24.23
480349	$Bi_{0.775}Sm_{0.225}O_{1.5}$	3.978	27.401	9.417056	24.38
480348	$Bi_{0.775}Nd_{0.225}O_{1.5}$	3.993	27.472	9.442118	24.41
410310	$Bi_{1.446}Nd_{0.554}O_3$	3.993	27.422	9.425959	24.45
480347	$Bi_{0.775}Pr_{0.225}O_{1.5}$	3.996	27.53	9.461289	24.37
410305	$Bi_{1.35}Pr_{0.65}O_3$	4.005	27.455	9.438319	24.49
480346	$Bi_{0.775}La_{0.225}O_{1.5}$	4.021	27.65	9.503597	24.42

tions from powder data is in general difficult. The total collection of patterns in the PDF will be released in the PDF-4/Full File RDB and all of the space groups listed in Table 2 will appear. Study of the differing trends in occurrences of classes of compounds, such as metals and alloys, in the range of space groups and prototype structures available may lead to a greater understanding of the factors affecting the determination of solid-state structure, and hence of physical properties, for materials. Entries from the PDF-4/Full File RDB and their associated literature references could form a reservoir of information on synthesis and properties of materials that crystallize in these space groups.

3. The $Bi_{2-x}A_xO_{3-\delta}$ system: a database study

Tailoring crystal structures and materials properties by preparing solid solutions is an important method in modern

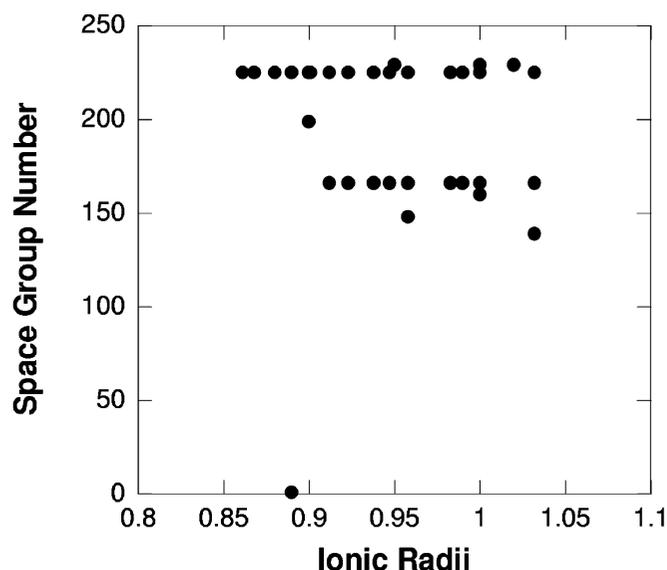**Figure 4**
Space groups in which $Bi_{2-x}A_xO_{3-\delta}$ can be stabilized: space group number *versus* ionic radius (Å) of the dopant A. (Note: space group number zero indicates that the space group was not reported for that entry.)

Figure 5
Space groups occurring in the $Bi_{2-x}A_xO_{3-\delta}$ system, where the ionic radius of A is between 0.8 and 1.1 Å. (Note: space group number zero indicates that the space group was not reported for that entry.)

Table 4
The results of various queries in the Release 1998 PDF database.

Hits	Search criteria
230	Two-element compounds
126	Cubic structures
64	Hexagonals/rhombohedrals
7	Orthorhombics
22	Monoclinics
4	Anorthics (triclinics)

materials design. Apart from routine phase identification, the PDF plays an important role in understanding the structural trends of these solid solutions. The advantages of a PDF-4 RDB in facilitating the study of solid solutions and their structural features are illustrated by taking the system $Bi_{2-x}A_xO_{3-\delta}$ (where A is a dopant at the bismuth site) as an example.

Bismuth sesquioxide (Bi_2O_3) exhibits a structural phase transition as a function of temperature and composition (Harwig, 1977, 1978; Harwig & Weenk, 1978). The high temperature δ phase of Bi_2O_3 has captivated the interests of materials scientists because of its superior oxide ion conductivity (Harwig & Weenk, 1978). It has been shown that the δ phase of Bi_2O_3 can be stabilized at room temperature by substituting rare earth or transition metal ions at the Bi site (Shuk *et al.*, 1996; Sillen, 1937; Schumb & Rittner, 1943). For the above reason, $Bi_{2-x}A_xO_{3-\delta}$ solid solutions have been studied extensively. In the present study, an attempt is made to comprehend $Bi_{2-x}A_xO_{3-\delta}$ solid solutions in terms of the dopants' ionic radii and oxidation states using the ICDD's PDF-4 RDB.

Fig. 4 shows the space groups in which $Bi_{2-x}A_xO_{3-\delta}$ can be stabilized against the ionic radii of the dopants. The δ phase of Bi_2O_3 has space group $Fm\bar{3}m$ (No. 225). The efforts to stabilize this phase by solid solution are reflected in the large number of $Fm\bar{3}m$ entries in the database. Interestingly, solid
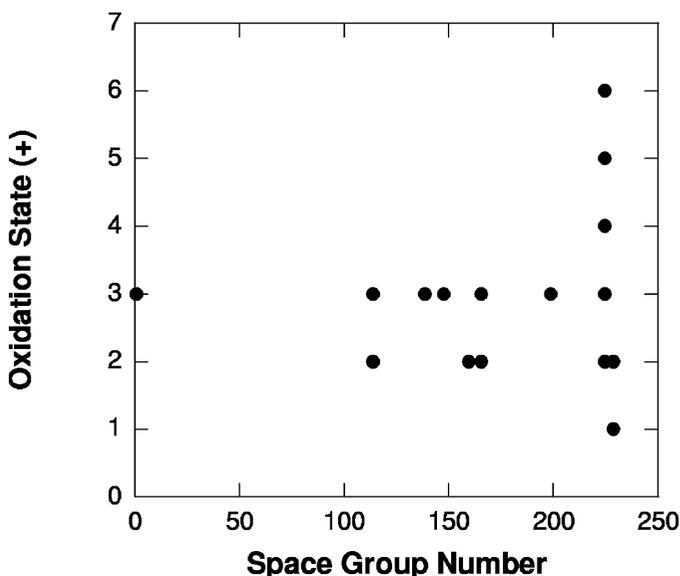
solutions with a dopant ionic radius between 0.85 and 1.05 Å (Fig. 5) crystallize predominantly in both $Fm\bar{3}m$ (No. 225) and $R\bar{3}m$ (No. 166). The factors that influence the occurrence of these two polymorphs in this region could be composition, temperature and preparation methods. It is worth noting that Bi^{3+} has an ionic radius of 0.96 Å, which is in the center of this region.

A comparison of the unit-cell dimensions of the $R\bar{3}m$ phases with that of the δ phase of Bi_2O_3 (a = 5.525 Å, space group $Fm\bar{3}m$) showed that the length of the rhombohedral cell edge is nearly equal to the body diagonal of the $Fm\bar{3}m$ unit cell ($3^{1/2}a_c$ = 9.5696 Å), as shown in Table 3. This group–subgroup
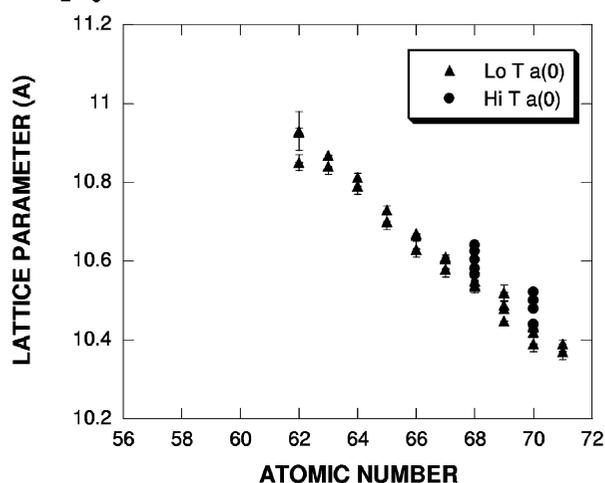

Figure 7
Ln_2O_3 lattice-parameter variation as a function of rare earth atomic number for compounds that crystallize in space group $Ia\bar{3}$.
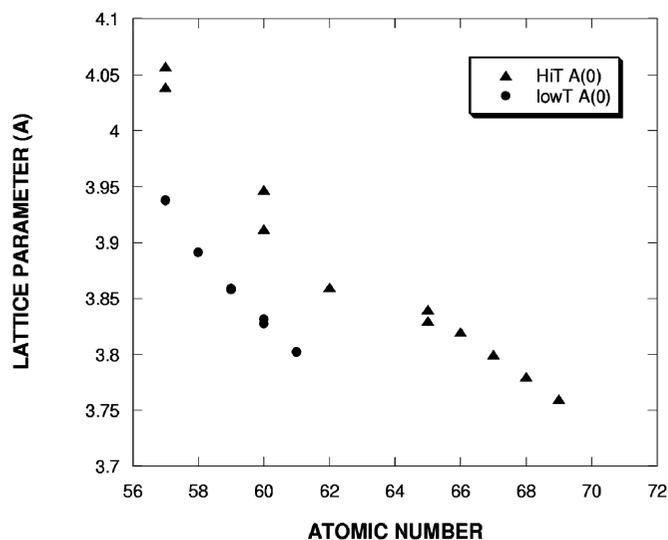

Figure 8
Ln_2O_3 lattice-parameter variation as a function of rare earth atomic number for compounds that crystallize in space group $P63/mmc$ and $P\bar{3}m1$. Only the a-axis lattice parameter is shown, but similar behavior is observed for the c axis.

Figure 6
Stabilization of $Bi_{2-x}A_xO_{3-\delta}$ in various space groups.

relationship ($R\bar{3}m$ is a non-isomorphous subgroup of $Fm\bar{3}m$) enables our understanding of such phase transitions.

The stabilization of $Bi_{2-x}A_xO_{3-\delta}$ in various space groups can also be analyzed in terms of the oxidation state of the dopant, as shown in Fig. 6. Substitution of di- and trivalent ions at the bismuth site crystallizes the structure in several space groups, including $Fm\bar{3}m$, indicating that by tuning composition one can stabilize $Bi_{2-x}A_xO_{3-\delta}$ in any one of these space groups. Tetra-, penta- and hexavalent ions crystallize $Bi_{2-x}A_xO_{3-\delta}$ in the $Fm\bar{3}m$ structure. Among all the oxidation states, monovalent dopants show a totally different behavior by crystallizing the compound in an AgI-type structure with space group $Im\bar{3}m$ (No. 229).

4. The Ln_2O_3 system: an heuristic data-mining example

The purpose of this investigation is to examine crystal symmetry and lattice-parameter variation over a wide range of rare earth ions that crystallize as rare earth sesquioxides. A series of queries were performed using the Release 1998 of the PDF. An RDB representation was employed to help facilitate the data-mining efforts. The first query contained results for two-element compounds with one rare earth constituent. The oxygen content was constrained to O_3. The search results were partitioned based on the published crystal symmetry. The results are reported in Table 4. Of the 126 cubic structure hits, 59 entries crystallize in the $Ia\bar{3}$ (No. 206) space group. One outlier entry was deleted from the analysis. The entries with space group $Ia\bar{3}$ (No. 206) are plotted in Fig. 7.

The high-temperature data in Fig. 7 were obtained for Er_2O_3 and Yb_2O_3 and represent a temperature-dependent series of measurements. From Fig. 7 we see that the light rare earths do not crystallize in the $Ia\bar{3}$ structure. Also, we note that the lattice contracts as a function of increasing atomic number. Fig. 8 shows that at low temperature a hexagonal structure is present for the light rare earths only. A different hexagonal phase is stable at high temperature across the entire rare earth series. Taken together, Figs. 7 and 8 provide crystal phase stability information on the rare earth sesquioxides. The rare earth contraction as a function of atomic number is observed

for all phases studied. This contraction arises from the filling of the inner $4d$ shells of the lanthanide ions as one traverses the rare earth series. It is important to note that these data were extracted from the PDF and as such demonstrate quite readily the high precision and accuracy of the experimental results reported here. Finally, there are over 75 literature references available for the cubic structure results listed above and a similar number available for the hexagonal structure materials. These are compiled in the RDB products discussed in this paper. The precision of the work undertaken by these investigators is truly remarkable.

5. Conclusions

Three examples have been presented to illustrate the power of data retrieval and mining capabilities using the PDF in PDF-4 RDB format. Empirically derived phase field stability diagrams can be easily developed. An attempt has been made to demonstrate mining techniques applied to RDB structural databases. The ICDD is developing a series of new RDB PDF products (PDF-4) that can be used for this purpose.

References

Faber, J. & Fawcett, T. (2002). *Acta Cryst.* B**58**, 325–332.
Faber, J., Kabekkodu, S. N. & Jenkins, R. (2001). *International Conference on Materials for Advanced Technologies*, Singapore.
Fink, W. L. (1962). *Scientific Information in the Fields of Crystallography and Solid State Physics*. Tokyo: Crystallographic Society of Japan.
Hanawalt, J. D., Rinn, H. W. & Frevel, L. (1938). *Ind. Eng. Chem. Anal. Ed.* **10**, 457–512.
Harwig, H. A. (1977). PhD thesis. State University Utrecht, The Netherlands.
Harwig, H. A. (1978). *Z. Anorg. Allg. Chem.* **444**, 151–166.
Harwig, H. A. & Weenk, J. W. (1978). *Z. Anorg. Allg. Chem.* **444**, 167–177.
Schumb, W. C. & Rittner, E. S. (1943). *J. Am. Chem. Soc.* **65**, 1055–1060.
Shuk, P., Wiemhofer, H. D., Guth, U., Gopel, W. & Greenblatt, M. (1996). *Solid State Ion.* **89**, 179–196.
Sillen, L. G. (1937). *Ark. Kemi Mineral. Geol.* **12A**, 1.