Inorganic Crystal Structure Database

ICSD

Scientific Manual

Version 1.1 (July 2008)

Contact: crysdata@fiz-karlsruhe.de

Contents

1 Introduction

This manual describes special information on all fields available in ICSD, independent of the retrieval software. The input rules for each field are the basis on which a search can be planned and performed.

At the moment two software packages are designed to meet the requirements of ICSD users. Special help features are included in these products. Download of subsets of ICSD records in well defined formats (*e. g.* CIF) and reuse specialised software is possible. In addition ICSD is available via STN International. (The data can be searched and displayed using a special retrieval language, special features like the display of structures or download of data are not supported.) FIZ Karlsruhe offers access to valuable specialised information but does not make any warranty as to the accuracy completeness, or usefulness. When feedback from customers reveals faults in the information, we shall make suitable efforts to correct the faults as quickly as possible.

The Scientific Manual is divided into three parts:

- **Section 2: "Input philosophy" of ICSD**
  Which structures are included in ICSD, and what are the criteria for the allocation of new ICSD entries?

- **Section 3: Field description, search strategies, and interpretation of the data**
  What are the search possibilities and restrictions that are due to the way the input was made? Why don't you find all compounds that contain Fe+2 by searching for "Fe+2" and how you should proceed to find them. It is described why there is often more than one citation record but only one set of authors etc.

- **Section 4: Description of algorithms for derived or calculated fields**
  Here a detailed description of all derived or calculated fields is given: Pearson symbol, Wyckoff sequence, minimal distances, ANX formula.

2 Scope and "Input philosophy" of ICSD

2.1 Chemical composition

ICSD started as an inorganic crystal structure database of published structures with atomic coordinates. The scope was gradually extended in the direction of intermetallic compounds with known atomic coordinates. Since 2003 FIZ Karlsruhe started to fill in the remaining gaps and will in future include all published intermetallic compounds. Until 2003 ICSD didn't contain any structures with C—H *or* C—C bonds ("organic compounds"). In 2003 this rule was modified: Now new entries should not contain both C—H *and* C—C bonds (*i. e.* compounds with tetramethylammonium will be included now as well as oxalates).

2.2 Structural information

The database contains fully determined structures with atomic coordinates. Coordinates for hydrogen atoms or vagabonding atoms like Na in Zeolites may be missing. Structures described as isotypic to known structures but without determination of free parameters are omitted. In some cases the lattice constants of other structure determinations or of isomorphic structures are given as a remark. If super- and substructures have been described, ICSD contains both of them as different entries.

2.3 "Input philosophy" – sources, checks, and evaluation

Most of the structures contained in ICSD are published in journals, only a few entries were submitted as private communications. Every completely determined structure corresponding to the rules cited above will be included into ICSD database. The entries are tested for formal errors, plausibility and logical consistency. The data are stored as published, they are not being standardised: Only some "exotic" space group settings will be transformed (*e. g. C* 1 1 2/*m* or *F* 1 $\overline{3}$ 2/*m*). As a rule the setting of the space group chosen by the author is considered as a valuable information that should not be changed. Since 2003 FIZ Karlsruhe has started to assign structure type classifications. In future updates this feature will enable the user to search for structurally closely related compounds.

A search in ICSD for a phase (defined by temperature, pressure and chemical composition) may yield in more than one entry. There are comments/remarks that give additional information on the quality of the data and refer to other related datasets of the same phase.

Please note that entries of “poor quality” will not be removed from the database. The final evaluation must be done by carefully studying the results of a search by the user. ICSD is a collection of datasets with the principal task to make this information available to the users.

3 Field description, search strategies, and interpretation of the data

3.1 Chemical name

This field contains the chemical name in English following IUPAC rules with oxidation state where necessary and corresponding to the first formula. This formula is calculated from all atoms with defined atomic coordinates. If undetermined atoms are included a remark is displayed in the atomic coordinates list. Oxidation state is given in Roman numbers. Phase information may be given after the hyphen. If two structures seem to be identical and have slight differences only due to special conditions details are given after a hyphen (*e. g.* doped material).

Examples:
Iron(III) hydroxide sulfate(VI)
Sodium carbonate – gamma
Aluminium oxide – Cr-doped

3.2 Mineral name

Due to a great effort by Prof. Allmann mineral names and group names are added to all records. New documents should contain such information. Details of origin are given after a hyphen. Tables with mineral names and group names are given in the special retrieval software.

Examples:
Albite high – from Rabb canyon, New Mexico,
Feldspar group
Cryolite – synthetic

Remember: It depends on the retrieval software how you may search for the names, group names, and additional information!

3.3 Chemical formula

The formula is coded as a structural formula and offers the opportunity to search for typical structure units (*e. g.* Si O4, Si2 O6) if a string search in the structural formula is performed. The sum formula is calculated from the structural formula and is stored in a separate field.

Remember: A string search in the structural formula will yield a result that is different from a search for the elements with subscripts in the sum formula. Note that the unit cell is not standardised. You must search for "Si O3" or *"Si2 O6" and you may miss "Si64 O128" if you only search for "Si O2"!*

The first formula must agree with the calculated formula from the structure determination (undetermined elements may be included with a remark in the atomic coordinates field). The second formula is the analytical formula or the complete formula for modulated structures. Atoms on identical sites should be combined by parentheses.

Examples:
K (H S O4)
((N H3)5 Co O2 Co (N H3)5) (N O3)5
Ca (H2 P O4)2 (H2 O)2
(Ni0.97 Co0.03) (Sb0.91 As0.08 Bi0.01)

3.4 Title of the publication

The Title of the publication is given in English, French or German.

3.5 Citation

The citation includes the year, the CODEN (with check-letter) for journals, volume, first page and last page.

Please note that a set of authors' names is only given for the first citation record.

There are several cases where more than one citation record occurs:

First Citation	Second Citation
Original publication	Publication contains data that have been used in the entry
Original publication	Cover-to cover journal
Full paper	Short Communication
Original paper	Paper that contains correction of the structure
Original publication	Publication with the same structure determination published in another journal

3.6 Authors

The surname is given followed by the initials of the first name. Some authors may be written differently depending on the language of the publication, *e. g.*, Gladyshevskii (in Russian) or Hladyshevsky (in Ukrainian). In a few cases in Chinese author names first or last name may be interchanged. Please inform us about any wrong writings.

3.7 Cell, *Z*, density

Information of unit cell parameters a, b, c, α, β, γ, measured density with standard deviations and the number of formula units (integer) is given; the measured density may be omitted if not given in the paper. All distances are in Ångstrøm.

3.8 Hermann–Mauguin Space Group Symbol

This symbol is used according to the International Tables for X-Ray Crystallography. If different origin choices are available, those space groups with origin at center are characterized by an additional "z", while an additional "s" is being used for special origins. Rhombohedral space groups with hexagonal cells are marked with an additional "h" ("hr" for reverse setting). Full symbols are used for monoclinic space groups to show the choice of the unique axis.

For all other space groups a list of coordinate triplets of a general position is given. Space groups that are not included in the International Tables need a critical evaluation. In a numerical database a space group "F13-2/m" may cause some trouble in statistical searches. It should be replaced by R3-mH (with 3/4 volume).

3.9 Atomic parameters

The element symbol is typed in upper and lower case. If the center of gravity has been determined only, one atom is put on this site, other atoms are treated like undetermined H

atoms in a remark. Atom identifiers in ICSD are principally running numbers. They may be different from the atom identifiers in the publication.

The oxidation state (formal charge) is given with sign (this number is a decimal with a maximum of 5 characters). If an oxidation state cannot be defined it is set to zero. This is the case for all intermetallic compounds. If electronic structure calculations (Hueckel band calculations, DOS, COOP) have been performed, an additional remark "ESC" (electronic structure calculations) is given. The total formal charge in the crystal must be zero.

Remember that it may be difficult to assign the correct formal charges to all elements. An average value may be used so that the total formal charge is zero. Do not search for exact oxidation states but for a range of oxidation states to include entries with average values.

3.10 Number of positions (multiplicity) and Wyckoff notation

Both are added for all sites. See also Wyckoff sequence in the third chapter. In the old input system "dummy" multiplicities were used for undetermined atoms, so the chemical formula was equal to the formula calculated from multiplicity, SOF and z. This may cause errors in some graphics programs.

3.11 Atomic coordinates, isotropic displacement parameters, site occupation, bonded H-atoms

The atomic coordinates *x*, *y*, *z* are given with a maximum of 7 characters each. The isotropic displacement (temperature) factor is given as "*B*" if defined by

$$\exp(-B\sin^2\theta/\lambda^2)$$

or "*U*" if defined by

$$\exp(-8\pi^2 U\sin^2\theta/\lambda^2) \quad .$$

Since 2003 isotropic and anisotropic temperature factors are given if both are available. Elder records contain either isotropic or anisotropic temperature factors (if anisotropic temperature factors are available). Displacement factors may have a maximum of 6 characters. The site occupation (max 5 characters) describes the statistical occupation of the site, *i. e.* the quotient of the actual amount of atoms and the maximum number of positions of the site. A special remark is generated for defect structures with SOF < 1. They need a special treatment when the Pearson symbol or the ANX formula is generated (see chapter 4). The number of undetermined H atoms bonded to another atom (*e. g.* H bonded to O in H_2O, or N in NH_4^+) is noted.

3.12 Anisotropic displacement parameters

Anisotropic displacement parameters are given in one of the following three different forms:

1. β (max 8 characters)

$$\exp\left(-(h^2\beta_{11} + k^2\beta_{22} + l^2\beta_{33} + 2hk\beta_{12} + 2hl\beta_{13} + 2kl\beta_{23})\right)$$

   with standard deviations if given.

2. B (max 8 characters each)

$$\exp\left(-\frac{1}{4}(h^2B_{11}a^{*2} + k^2B_{22}b^{*2} + l^2B_{33}c^{*2} + hkB_{12}a^*b^* + klB_{23}b^*c^* + hlB_{13}a^*c^*)\right)$$

   with standard deviations if given.

3. U (max 8 characters each)

$$\exp\left(-2\pi^2(h^2U_{11}a^{*2} + k^2U_{22}b^{*2} + l^2U_{33}c^{*2} + 2hkU_{12}a^*b^* + 2hlU_{13}a^*c^* + 2klU_{23}b^*c^*)\right)$$

   with standard deviations if given.

3.13 Remarks

There are standard remarks, free text remarks, and test remarks that are messages from the test program.

The standard remarks consist of three characters. They are searchable and listed below. Please note that the remarks TEM (temperature in K) and PRE (pressure in MPa) contain numerical values and will be made numerically searchable in future updates.

All remarks will be displayed in a full sentence.

Additional information is also given if available:

POL	contains additional type information
DIS	may contain a description of the disorder
PDF	contains the PDF number.

Standard remarks and the related full text that may be completed by additional information if available:

ABC	Absolute configuration given
AHT	Anharmonic temperature factors given
CGD	The structure contains atom groups whose centre of gravity only has been determined
COA	Published data has been corrected through correspondence with the author
COR	This publication corrects errors in an earlier one
DEF	Defect structure. At least one site occupation is less than one.
DIS	Disordered structure that cannot adequately be described by numerical parameters
EDP	Electron diffraction (powder)
EDS	Electron diffraction (single crystal)
EMP	Electron microscopy of a single powder particle
ESC	Electronic structure calculations
MAG	Magnetic structure also determined
MIN	Compound with mineral name
MOD	Modulated structure or misfit-layer structure
NDP	Neutron diffraction (powder)
NDS	Neutron diffraction (single crystal)
NMR	NMR spectroscopy data given
ODS	Order-disorder structure
PDF	The structure has been assigned a PDF number
POL	Polytype structure
PRC	Preliminary publication
PRE	Pressure in MPa
REF	New refinement based on previously measured intensities
RVP	Rietveld profile refinement applied
SFP	Structure determined from projections
SNP	Synchrotron radiation (powder)
SNS	Synchrotron radiation (single crystal)
TEM	Temperature in Kelvin
THE	Structure calculated theoretically
TWI	Structure determined on a twinned crystal
XDP	X-ray diffraction (powder)

In addition to the standard remarks there is a free text field for additional comments (*e. g.* cells of isomorphic compounds if given in the paper).

3.14 Test Codes and Related Remarks in ICSD

Test codes are messages that originate from the test program used during the creation of an entry into the database. The appearance of these test codes depends on the setting of tolerances in the program. A discussion is going on whether these test messages should be used for internal purposes only and should not be displayed by the retrieval software. Especially 76 is critical. This message appears too often with the present setting of tolerances.

21	Difference between the calculated formula and the formula in formula-field is tolerable.
22	Deviation of the charge sum from zero tolerable.
23	Calculated density unusual but tolerable.
51	No *R* value given in paper.
52	Temperature factor implausible but agrees with the paper.
53	At least one temperature factor missing in the paper.
54	A site occupation is implausible but agrees with the paper.
55	Lattice parameters are unusual but agree with the paper.
56	Reported temperature factors are non-standard (not included in entry).
74	Coordinates are those given in the paper but are probably wrong.
75	Reported coordinates contain an error. Values corrected.
76	Interatomic distances appear to be too short.

3.15 *R*-Index

The *R*-index contains up to five characters

$$R = \frac{\sum \Delta F}{\sum F} \quad ,$$

if given; otherwise the lowest *R*-value is used (R_p for Rietveld refinement if possible, R_{Bragg} may be given as an additional remark).

4 Description of algorithms for derived or calculated fields

4.1 ANX Formula

The ANX formula is generated according to the following rules:

- H^+ is not taken into account, even if coordinates are available.
- The coordinates for all sites of all other atoms must be determined.
- Different atom types on the same position (*e. g.* for solid solutions) are being treated as one single atom type. The relevant atom type is the one with the highest site occupation factor. If the SOFs are equal the first atom type is the relevant atom type.
- Exception to this rule: if anions and cations occupy the same site they will not be treated as one atom type.
- All sites occupied by the same atom type are combined unless the oxidation state is different.
  $Fe^{2+}(Fe^{3+})_2O_4 \longrightarrow AB_2X_4$
  $(Fe^{2.6667+})_3O_4 \longrightarrow A_3X_4$
- For each atom type the multiplicities are multiplied by the SOFs and the products are added. The sums are rounded and divided by the greatest common divisor.
- If the rounded sum is equal to zero all sums are being multiplied by a common factor so that the smallest sum is equal to 1.0, so no element will be omitted.
- Cations are assigned the symbols A–M, neutral atoms N–R and anions are assigned X, Y, Z, S–W.
- The symbols are being sorted alphabetically and the characters are assigned according to ascending indices: AB_2X_4, not A_2BX_4.
- All ANX formulae with more than 4 cation symbols, 3 neutral symbols or 3 anion symbols are deleted. This measure limits the number of different ANX formulae.

Example: Garnet group

Chemical formula	ANX formula
$Mg_3Al_2(SiO_4)_3$	$A_2B_3C_3X_{12}$
$Ca_3(Al_{1.3325}Fe_{0.6675})Si_3O_{12}$	$A_2B_3C_3X_{12}$
$(Mg_{2.7}Fe_{0.3})(Al_{1.7}Cr_{0.3})Si_3O_{12}$	$A_2B_3C_3X_{12}$

4.2 Pearson symbols

Three datasets are required to create a Pearson symbol:

1. Symbol for the crystal system
   - *a* = triclinic (anorthic)

- *m* = monoclinic
- *o* = orthorhombic
- *c* = cubic
- *h* = trigonal
- *h* = hexagonal
- *t* = tetragonal

2. Symbol for Bravais-Type
   - *P* = primitive
   - *A* = *A*-face centered
   - *B* = *B*-face centered
   - *C* = *C*-face centered
   - *F* = face centered
   - *I* = body centered
   - *R* = rhombohedral

   In future the symbols *A*, *B*, *C* will be given as "*S*" (one side centered) for monoclinic and orthorhombic space groups (monoclinic also *I*) so the Pearson Symbol will become independent of the cell transformations.

3. The number of atoms in the unit cell
   This number is being calculated from the sum of all multiplicities and the respective site occupation factors for each atom type. The data are rounded. In rhombohedral space groups the number of atoms depends on the trigonal or hexagonal setting of the space group. In ICSD the number of atoms in hexagonal settings of rhombohedral space groups should be divided by 3 to match the data from Pearson's Handbook and to get numbers invariant to cell transformations.

4.3 Wyckoff Sequence

Sites with SOF < 1 are treated in analogy to the treatment for the ANX formula (see above). Symmetrically equivalent sites (represented by a Wyckoff letter) are counted and the number is given along with the Wyckoff letter.

The Wyckoff letters are sorted alphabetically (descending sort) and written along with their number, each new Wyckoff letter is separated by a blank.

Example: *g*10 *f*8 *e*4 *b*2 *a*6

This sequence is not unique, but depends in some space groups on the choice of origin. *E. g.* for spinels the sequence changes from "*f d a*" to "*f c b*" after an origin shift of 1/2 1/2 1/2.

4.4 Minimum Distances

For the calculation of minimum distances a program from the Cambridge Crystallographic Subroutine Library (CCSL) is used. Detailed information is given in the internet:

`http://www.ill.eu/sites/ccsl/html/ccsldoc.html`

