Composition: A Critical Property for Chemical and Material Databases

Charles W. Moulton

Chemical Abstracts Service, Columbus, Ohio 43210

Received June 4, 1992

Knowledge of precise compositions is crucial for applications of materials. Traditional representations of composition do not at times adequately describe certain types of materials of interest to users of the numeric property databases. Changes in registration practices proved to be necessary so that the compositions of some of these materials, as recorded in the CA Registry File, could be linked to the pertinent property data in the numeric files. Enhanced registration practices adopted for nonstoichiometric compounds, solid solutions, and alloys with standardized compositions are described.

I. INTRODUCTION

It might seem to belabor the obvious to discuss the subject of composition in a symposium that is concerned with chemical and material property databases. To obtain meaningful property data for particular substances, those substances must be specified as precisely as possible. Substance identification takes many forms—chemical name, class or family designation, trade name, formula, structure, etc. What will mainly be discussed in this paper is how certain types of substances of interest to users of the numeric property files are described in the CAS Registry File and how the Registry File relates to these numeric property files.

II. THE REGISTRY FILE

The CAS Chemical Registry began operation over 25 years ago as an internal production tool to support indexing, a function that it still serves. In more recent years it has become established as a major source for the identification of chemical substances as well. Registry Numbers assigned to particular substances are primary identifiers for many purposes: they serve as search terms in many of the online files that are concerned with materials, and they are required by various regulatory agencies that control the manufacture, sale, transportation, usage, and disposal of materials of commerce.

The CAS Registry File is a database that contains various types of information on substances: chemical names, elemental symbols and chemical formulas, structural representations, and other means of compositional descriptions. Substances appearing in this file include the elements and their ionic and isotopic forms, inorganic and organic compounds, polymers, biomolecular sequence fragments, alloys and cermets, mixtures and solid solutions, and the particles of interest to nuclear chemists. Substance descriptions range from precisely defined compositions and unambiguous structures, through those that are only partially defined, to identifications provided only as trade names for commercial products or by the designations given by physicists for the subatomic particles. Well over 11 million unique substances now appear in this file.[1]

The CAS Registry File is central to many of the databases that make up the property data networks. The Registry File provides numerical identifiers, the Registry Numbers, for particular substances. These numbers are for the most part linked to definite compositions and to chemical structures, but also to less precisely defined substances, trade names, etc. They are associated not only with the CA *index names* but also with other names that appear in the literature—what are referred to as *synonyms*. While it is the index names that are published in the CA Subject Indexes, the Registry Numbers themselves are identifiers that are not bound by the conventions and vagaries of any one nomenclature system. Index names can change from time to time, but the numbers assigned to the substances do not.

Registry Numbers appear in the six files that make up the Chemical Property Data Network (CPDN) and also in most of the other numeric files available on STN. When a search is carried out by substance name in the CPDN databases by means of the menu mode, the Registry File is first queried behind the scenes. When the name is located, it is the Registry Number that then becomes the search term in the property data files, a mechanism that is transparent to the searcher. This feature is further discussed in another paper presented in this symposium.[2]

III. ORGANIC COMPOUNDS

Little attention is paid here to the subject of organic substances. But by way of illustration, the display from the Registry record for a simple hydrocarbon, butane, is given in Figure 1.[3] The Registry Number appears in the RN field, the various names in the CN fields (with the CA index name appearing first and the various synonyms following), and the chemical formula in the MF field. The Registry Number Locator field (LC) identifies those STN files that contain information on butane (i.e., those to which the RN is posted); among the numeric files cited here are Beilstein, DIPPR, HODOC, and SPECINFO.

Figure 2 is a portion of the record for a polymer, nylon 66. The full record in this case contains over 200 synonyms, by far the greatest number being trade names. The Class Identifier (CI) field illustrated here provides the means to segment the Registry File into broad classes of substances; the code PMS identifies a polymer.

As an example of how a Registry record is linked to one of the numeric files, Figure 3 shows the display from the Plastics Materials Selection Database (PLASPEC) for a form of nylon 66. This record identifies it as a thermoplastic of the nylon family and lists the physical properties reported for a particular grade produced by DuPont. The PLASPEC database contains over 900 separate records for properties of nylon 66 supplied by various manufacturers, with different trade names, in different physical forms, with various treatment conditions, etc.

IV. INORGANIC COMPOUNDS

Typical of the Registry record for many simple inorganic compounds is that for the binary compound uranium dioxide (Figure 4). The CA index name is *uranium oxide*, accom-

0095-2338/93/1633-0027$04.00/0

```
RN   106-97-8
CN   Butane  (CA INDEX NAME)
CN   n-Butane
CN   Diethyl
CN   Liquefied petroleum gas
CN   LPG
CN   R 600
MF   C4 H10
LC   ANABSTR, APILIT, ... BEILSTEIN, BIOSIS, CA,
     CAOLD, CASREACT, CHEMLIST, ... DIPPR, DSL,
     EINECS, EMBASE, GMELIN, HODOC, IFICDB, IFIPAT,
     ... NISTFLUIDS, ... SPECINFO, TRCTHERMO, TSCA
```

$H_3C\text{-}CH_2\text{-}CH_2\text{-}CH_3$

Figure 1. Registry record for butane.

```
RN   32131-17-2
CN   Poly[imino(1,6-dioxo-1,6-hexanediyl)imino-
       1,6-hexanediyl]  (CA INDEX NAME)
CN   Nylon 66
CN   Zytel 101
CN   Poly(hexamethyleneadipamide)
CN   Adipic acid-hexamethylenediamine polymer, SRU
CN   Amilan CM 3001
CN   Zytel 42
CN   Maranyl A 100
CN   Polypenco 66D
CN   Poly(hexamethylenediamine-adipate)
CN   Polyamide 66
CN   Hexamethyleneadipamide polymer
CN   Ultramid A
CN   Torayca N 66
CN   Wellamide 6600
MF   (C12 H22 N2 O2)n
CI   PMS
LC   ANABSTR, BIOSIS, CA, ...
     PDLCOM, PLASPEC, TSCA
```

Figure 2. Registry record for nylon 66.

```
Database: PLASPEC/Manufacturer's Data    AN: 400-11281


Material:               Thermoplastic
Generic Family:         Nylon
Chemical Type:          Nylon 66
Supplier:               Du Pont, Polymer Products Dept.
Process Type:           Injection molding
Trade Name and Grade:   Zytel 101    F NC-10
CAS Registry Number:    32131-17-2 (Nylon 66)
Features:               High-speed processing; Lubricated
Price:                  254 cents/lb; 10.45 cents/in**3

     Processing/Physical Characteristics

Density:                            1.14 g/cm**3

     Mechanical Properties

Tensile Strength at Yield:          1.2E+04 psi
Elongation at Break:                52%
Modulus of Elasticity
        Flexural:                   4E+05 psi
Hardness
        Rockwell:                   R121
Izod Impact Energy, Normalized:     1 ft*lbf/in for 0.125
                                    in wide specimen

     Thermal Properties

Heat Distortion Temperature:        455 deg F at 66 psi;
                                    194 deg F at 264 psi
Melting Point:                      491 deg F
Flammability
               UL Rating:            V-2
           Thickness:               0.25 in
```

Figure 3. PLASPEC record for nylon 66.

panied by the parenthetical line formula UO_2. The molecular formula in the MF field appears in the usual Hill format (alphabetical by element symbol, except when carbon is present).[1]

The resurgence of research in solid-state chemistry in recent years has led to the occurrence of increasingly complex compositional representations for inorganic substances in the literature. Several years ago, the need for more precise means of description of the compositions of nonstoichiometric compounds and solid solutions was recognized at CAS. Registration practices in the area of solid-state inorganic compounds were enhanced so that the more complex representations could be adequately indexed.

```
RN   1344-57-6
CN   Uranium oxide (UO2)  (CA INDEX NAME)
CN   C.I. 77915
CN   Uranium dioxide
CN   Urania
CN   γ-Uranium dioxide
CN   Urania (UO2)
CN   Uranium oxide
CN   Uranium(4+) dioxide
MF   O2 U
LC   ANABSTR, BIOSIS, CA, CAOLD, CHEMLIST,
     ... NISTTHERMO, TSCA, VTB
```

Figure 4. Registry record for uranium dioxide.

```
   Current Registration                      Pre-12CI Registration

RN   136140-70-0                         RN   1344-57-6D
CN   Uranium oxide (UO2.14)              CN   Uranium oxide (UO2)
MF   O . U                                      (oxygen-excess)
AF   O2.14 U
CI   TIS


RN   136140-69-7                         RN   1344-57-6D
CN   Uranium oxide (U0.99O2.14)          CN   Uranium oxide (UO2)
MF   O . U                                      (nonstoichiometric)
AF   O2.14 U0.99
CI   TIS
```

Figure 5. Registry records for nonstoichiometric uranium dioxides.

Nonstoichiometric compounds are those that do not obey the elementary laws of definite composition and of multiple proportions, i.e., they cannot be simply represented by traditional chemical formulas that employ only simple integers as coefficients.[4] Although these so-called berthollides have been known since the early years of the 19th century, the indexing of their actual compositions was not stressed until recently.

With the beginning of the 12th Collective Period in 1987, enhancements were made to registration practices to allow for formulas with coefficients in the form of decimal fractions. Elemental ratios for nonstoichiometric compounds are now recorded as they are reported in the original literature (to the limit of two decimal places).

The current Registry records for two nonstoichiometric forms of uranium dioxide are contrasted in Figure 5 with the means of registration used prior to the 12th Collective Period. The first of these compositions has an excess of oxygen, and the second has a similar oxygen excess and, additionally, a slight deficiency in the uranium content. The index name in each case is *uranium oxide* as before, with the nonstoichiometry indicated in the parenthetical line formulas.

Since it is not feasible to derive structural connection tables for substances of this type, the current means of registration is in the form of tables that express the elemental compositions. This mode of registration is identified by the code TIS (for Tabular Inorganic Substance) in the CI field. What appears in the MF field is a simple alphabetical listing of the elemental components, with the Hill formula now appearing in the Alternate Molecular Formula (AF) field.

Previously, such compositions were indexed as uranium dioxide itself, with information pertaining to deviations from stoichiometry conveyed by phrases such as "oxygen-excess", "uranium-deficient", or "nonstoichiometric" in the index entries. These phrases did not, however, appear as parts of the Registry records. (The suffix D appended to the Registry Number denotes the inclusion of derivative information in the index entries, i.e., that the basic composition has been in some way modified.)

Akin to nonstoichiometric compounds are the *solid solutions* or mixed crystals. Formulas with integral coefficients all too frequently do not suffice to describe such compositions. The next three figures illustrate the Registry records for several electronic materials that commonly exist as solid solutions.

```
  Current Registration                     Pre-12CI Registration

RN  137852-39-2                          RN  12060-00-3D
CN  Lead titanium zirconium oxide        CN  Lead titanium oxide (PbTiO3)
      (PbTi0.1-0.8Zr0.2-0.9O3)                 (solid solns. with lead zirconate)
MF  O . Pb . Ti . Zr
AF  O3 Pb Ti0.1-0.8 Zr0.2-0.9            RN  12060-01-4D
CI  TIS                                  CN  Lead zirconium oxide (PbZrO3)
                                               (solid solns. with lead titanate)
```

Figure 6. Registry record for a lead zirconate titanate ferroelectric ceramic.

```
  Current Registration                     Pre-12CI Registration

RN  137750-81-3                          RN  1303-00-0D
CN  Gallium indium arsenide phosphide    CN  Gallium arsenide
    (Ga0.11In0.89As0.25P0.75)            RN  12063-98-8D
MF  As . Ga . In . P                     CN  Gallium phosphide
AF  As0.25 Ga0.11 In0.89 P0.75           RN  1303-11-3D
CI  TIS                                  CN  Indium arsenide
                                         RN  22398-80-7D
                                         CN  Indium phosphide
                                               (solid solns. with
                                               Group IIIA pnictides)
```

Figure 7. Registry record for a gallium indium arsenide phosphide semiconductor.

```
RN  119855-33-3
CN  Barium calcium copper strontium ytterbium oxide
    (Ba0.7Ca0.6Cu3Sr0.7YbO6-8)
MF  Ba . Ca . Cu . O . Sr . Yb
AF  Ba0.7 Ca0.6 Cu3 O6-8 Sr0.7 Yb
CI  TIS


RN  136255-18-0
CN  Barium calcium copper carbonate oxide
      (Ba4CaCu2.24(CO3)0.5O6.96)
MF  C O3 . Ba . Ca . Cu . O
AF  C0.5 Ba4 Ca Cu2.24 O8.46
CI  TIS
```

Figure 8. Registry records for cuprate superconductors.

The first of these (Figure 6) is for a ferroelectric ceramic, a lead zirconate titanate (PZT), which can be regarded as a solid solution of lead titanate and lead zirconate. The CA index name is that for a mixed-metal oxide, *lead titanium zirconium oxide*, accompanied by a parenthetical line formula stating the compositional ranges of the titanium and zirconium. Prior to the 12th CI, each end member of the pseudobinary system (i.e., lead titanium oxide and lead zirconium oxide) was separately indexed, with phrases of the type "solid solns. with ..." included in the index entries. As with the nonstoichiometric compounds, these phrases did not appear in the Registry records, but the suffix D was added to the Registry Numbers to indicate chemical modification.

A composition in the III–V semiconductor system gallium indium arsenide phosphide is shown in Figure 7. The current means of registration uses a similar solid-solution approach. Formerly the substance would have been broken apart into the four presumed end members of the system; phrases such as "solid solns. with Group IIIA pnictides" appeared in the index entries but, again, were not included within the Registry records.

The complex compositions of two of the high-temperature superconducting cuprates are illustrated in Figure 8. For the first of these, an alkaline rare earth cuprate, the oxygen content is recorded as a range, while the other elemental compositions are precisely stated, a common feature of these systems.[5,6]

A further level of complexity is shown by the second example, a carbonate-containing cuprate, which represents the situation of nonstoichiometry within a double-anion solid-solution system. The carbonate ion is treated here as one of the constituents in the tabular representation and appears as a unit in the MF field.

For these complex cuprates, pre-12th CI registration practices frequently provided no suitable means for retaining compositional information beyond identification of the elemental components in the names. There might be no practicable way to identify completely the particular stoichiometric forms from which the compositions deviated or to break them apart into a reasonable group of supposed end members. The significant physical parameters that relate to superconductivity, principally the transition temperature, are extraordinarily sensitive to slight differences in composition. These systems provide striking examples of the need for specifying the compositions of these types of inorganic substances to levels of detail well beyond what would have been imagined just a relatively short time ago.[7]

```
RN  81796-60-3
CN  Aluminum alloy, base, Al 94,
      Cu 3,Li 2,Mg 1,Zr 0.1
MF  Al . Cu . Li . Mg . Zr
CI  AYS

Component  Component
             Percent
=========+=============
   Al         94
   Cu         3
   Li         2
   Mg         1
   Zr         0.1


RN  12616-53-4
CN  Aluminum alloy, base,
      Al 93,Al2O3 7
MF  Al2 O3 . Al
CI  AYS

Component  Component
             Percent
==========+==========
   Al         93
   Al2O3      7
```

Figure 9. Registry records for an aluminum alloy and an aluminum–alumina cermet.

V. ALLOYS

Registration of alloy compositions began soon after the Registry System was instituted. Since neither chemical formulas nor structures in the traditional sense are meaningful, alloy compositions have been registered in terms of their elemental components and weight percentages, according to standard metallurgical usage. The single element present in largest amount is labeled the *base*. All other (*nonbase*) elements are recorded down to 0.1 wt %, but after rounding, only two significant figures at most are retained and trailing zeros are dropped.

This means of registration is illustrated for an aluminum alloy in Figure 9. The index name is *aluminum alloy, base*, followed by a linear listing of the constituent elements and their weight percentages. The MF field contains an alphabetical listing of the elemental symbols. The code AYS in the CI field identifies the alloy mode of registration. In addition to the compositional information included as part of the index names, tabular listings of the compositions appear in the Registry records for all alloys for which there are specified numerical compositions.

Cermets are registered by the same practices that apply to the all-metal alloys and are given alloy-like index names, as shown for an aluminum–alumina cermet.

The registration practices adopted many years ago for alloys had the limitation that closely related compositions could be merged into single Registry records. For the numeric data files, the consequence of this would have been that significant property distinctions related to small compositional differences would have been suppressed. Registration practices for those alloys for which there are defined standards have been modified to allow for greater levels of precision. For about the last 3

```
RN   137536-93-7
CN   Nickel alloy, base, Ni 50-64,Mo 15.0-17.0,
     Cr 14.50-16.50,Fe 4.0-7.0,W 3.00-4.50,Co 0-2.50,
     Mn 0-1.0,Si 0-1.0,Cu 0-0.50,C 0-0.10,P 0-0.040,
       S 0-0.030 (UNS W80002)  (CA INDEX NAME)
CN   Hastelloy C electrode
CN   UNS W80002
CN   AWS ENiCrMo-5
CN   MIL-3N1C
CN   ENiCrMo-5
MF   C . Co . Cr . Cu . Fe . Mn . Mo . Ni . P
      . S . Si . W
CI   AYS
```

Component	Component Percent
Ni	50 - 64
Mo	15.0 - 17.0
Cr	14.50 - 16.50
Fe	4.0 - 7.0
W	3.00 - 4.50
Co	0 - 2.50
Mn	0 - 1.0
Si	0 - 1.0
Cu	0 - 0.50
C	0 - 0.10
P	0 - 0.040
S	0 - 0.030

Figure 10. Registry record for a Hastelloy.

```
RN   11107-04-3
CN   Iron alloy, base, Fe 62-72,Cr 16.00-18.00,
     Ni 10.00-14.00,Mo 2.00-3.00,Mn 0-2.00,
     Si 0-1.00,C 0-0.08,P 0-0.045,S 0-0.030
      (UNS S31600)  (CA INDEX NAME)
CN   AISI 316
CN   RM 30
CN   DIN 1.4401
CN   SUS 32
CN   Thermalloy A28M
CN   EI 401
CN   Cr18Ni10Mo2
CN   Inox 316
CN   Stainless steel 316NG
MF   C . Cr . Fe . Mn . Mo . Ni . P . S . Si
CI   AYS
```

Figure 11. Registry record for stainless steel 316.

years, alloys with standard defined compositions have been registered to the levels of detail specified by the standardizing bodies [e.g., the American Iron and Steel Institute (AISI) or the Copper Development Association (CDA)]. As many significant figures and decimal places are retained as are required to completely describe these compositions.

A typical example of this enhanced alloy registration appears in Figure 10 for a Hastelloy. Weight percentages extend to four significant figures, and the minor elemental constituents (in this case phosphorus and sulfur) that would previously have been below the threshold for inclusion are retained. The UNS (Unified Numbering System) designation is included in the index name as a further identifier.

Another example, for stainless steel 316, is provided in Figure 11. The index name is *iron alloy, base*, followed by the composition in a linear listing and the parenthetical AISI designation. The full record contains nearly 50 synonyms for this stainless steel. (The CA index name *stainless steel* is used when numerical compositions are not known.)

An example of a steel alloy for which the CA index name is *steel* is 4142H (Figure 12). The standard designation AISI 4142H here appears in the CN field as the compositional identifier. The complete numerical composition is given in tabular form as part of the Registry record. The STEELTUF file, as cited in the LC field, is the source of property data on this steel.

The impetus for these enhancements for alloy compositions was the appearance of the numeric property databases on STN. Greater precision was imperative so that alloys with standard compositions could be directly related to the property information. To date, however, these enhancements have been

```
RN   125352-82-1
CN   Steel, (AISI 4142H)  (CA INDEX NAME)
CN   A540-B22
CN   AISI 4142H
CN   708H42
CN   UNS H41420
CN   ASME SA540-B22
MF   C . Cr . Fe . Mn . Mo . P . S . Si
CI   AYS
LC   STEELTUF
```

Component	Component Percent
Fe	97 - 98
Cr	0.75 - 1.20
Mn	0.65 - 1.10
C	0.39 - 0.46
Si	0.15 - 0.35
Mo	0.15 - 0.25
S	0 - 0.040
P	0 - 0.035

Figure 12. Registry record for a carbon steel.

applied only to those alloys for which there are defined standards.

VI. SUMMARY

Only some of the recent changes in registration practices have been described here. A major point of this discussion is that the CAS Registry File is a dynamic file, periodically undergoing upgrading as the needs for greater precision in the recording of compositional information and for improved search capabilities are recognized. This might result from new directions that appear in the literature or from needs that are expressed by users.

Linking of the compositions of materials to physical data in the numeric files extends the concept of a chemical registry. As was described, registration practices not only relate to how chemicals and materials are identified in the numeric files, but can in turn be affected by the availability of those databases.

Some of the areas in which materials of potential interest for users of the numeric files are not now completely identified within the Registry File are the following:

crystallographically distinct polymorphic forms (with the exception of minerals)

phase contents of multiphasic ceramics

composite materials, such as laminates, reinforced plastics, etc. (other than the cermets)

The CAS Registry System at its inception was based primarily on the chemical compositions and structures of specific compounds; this is still largely the case. The Registry File is pivotal to the identification of compositions in the numeric property databases, so that as such databases expand into the areas of composite and multiphasic materials, the need for further changes to registration practices will likely arise.

REFERENCES AND NOTES

(1) For a more complete description of the content of and search capabilities in the CAS Registry File, see *The Registry File Database Description*, Chemical Abstracts Service, American Chemical Society: Columbus, OH, 1991.

(2) Drago, V. J.; Kaufman, J. G. Chemical and Material Property Data Networks on STN International. *J. Chem. Inf. Comput. Sci.*, in this issue.

(3) For this and later examples of Registry records, various features not pertinent to the particular discussion have been omitted for clarity. Displays in some cases have been reformatted to facilitate publication.

(4) Hannay, N. B., Ed. *Treatise on Solid State Chemistry*; Plenum Press: New York, 1973–1976.

(5) Nelson, David L.; Whittingham, M. Stanley; George, Thomas F.; Eds. *Chemistry of High-Temperature Superconductors*; ACS Symposium Series 351; American Chemical Society: Washington, D.C., 1987.

(6) Mason, T. O. Defect Chemistry of High T_c Superconducting Cuprates. *Key Eng. Mater.* **1992**, *66 & 67*, 503–536.

(7) Dess, Howard M. Online Search Strategies for Semiconductor or Superconductor Materials. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 84–89.