An XML Namespace for IUPAC

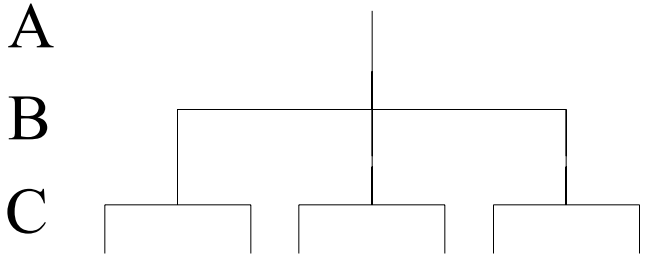
Jeremy Frey, Robert Lancashire, Alan McNaught,
Peter Murray-Rust, Henry Rzepa, Steve Stein

(and a dozen more)

CAS/IUPAC Conference on Chemical Identifiers and XML for Chemistry
Columbus, OH
July 1, 2002

IUPAC	XML
Standards for Chemical Communication	Standards for Representation
Terminology	Ontology
Data Identification	Tags
Standard Data	Instances

XML – Tags and Tools



Data-Type Tree

tags

<A><B><C>data</C></B></A>

XML Serialization

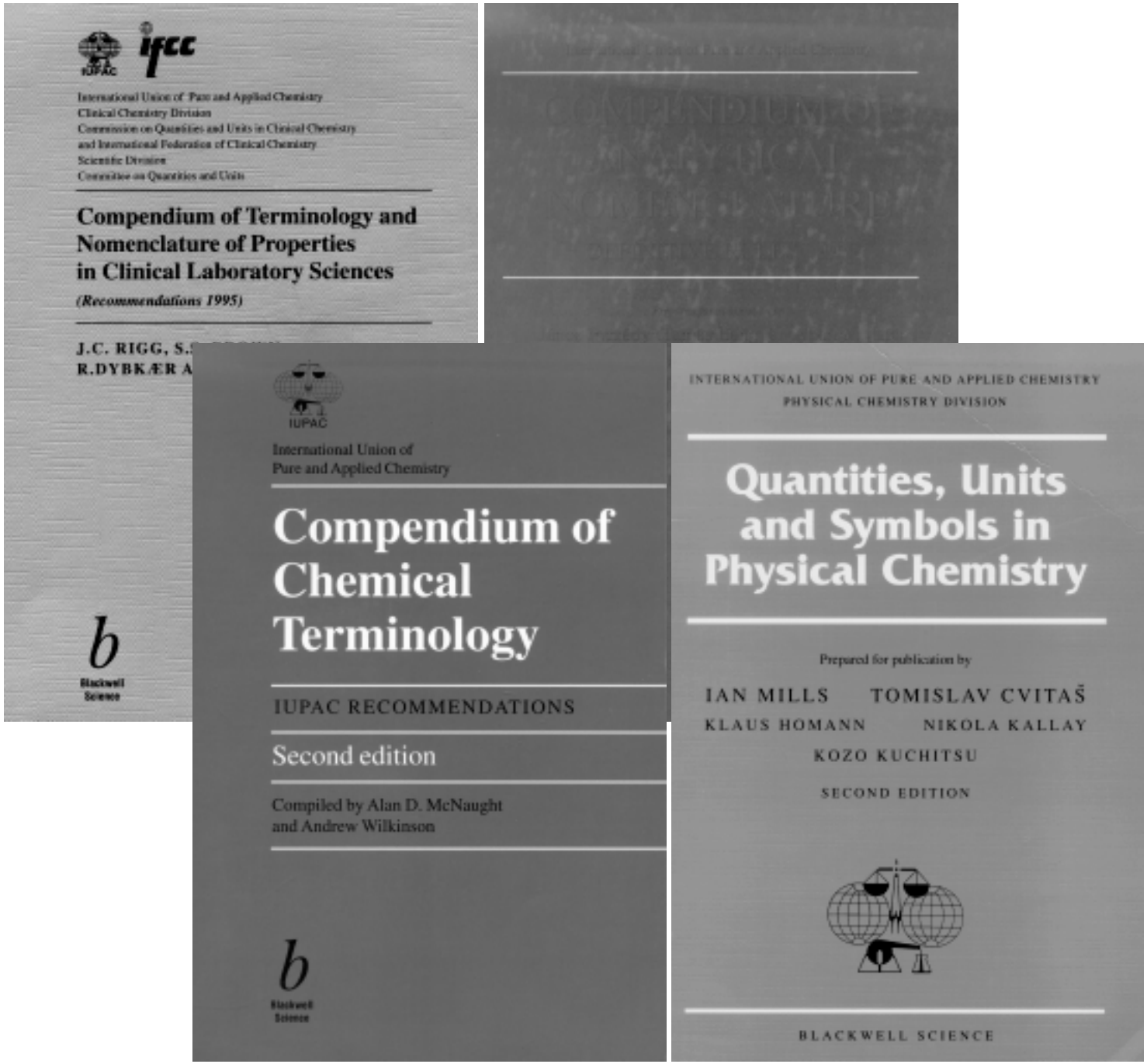
tools

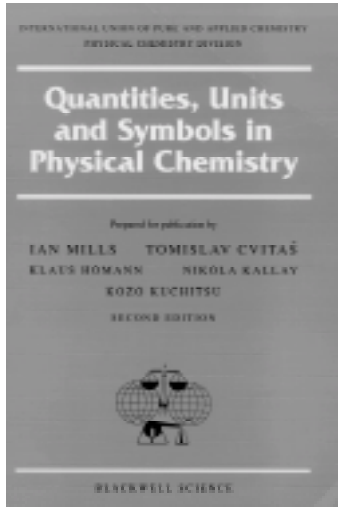
Chemist, Experiment,
Database

IUPAC ‘Namespace’

- Nomenclature
- Glossaries
  - Diverse, broad to specialized
- Data Representation
- Chemical Data
  - Periodic Table (relative molar mass)
  - Reference Data

Many in cooperation with other international standards bodies


IUPAC
ifcc
International Union of Pure and Applied Chemistry
Clinical Chemistry Division
Commission on Quantities and Units in Clinical Chemistry
and International Federation of Clinical Chemistry
Scientific Division
Committee on Quantities and Units
Compendium of Terminology and Nomenclature of Properties in Clinical Laboratory Sciences
(Recommendations 1995)
J.C. RIGG, S.
R.DYBKÆR A
b
Blackwell Science
IUPAC
International Union of Pure and Applied Chemistry
Compendium of Chemical Terminology
IUPAC RECOMMENDATIONS
Second edition
Compiled by Alan D. McNaught and Andrew Wilkinson
b
Blackwell Science
INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY
PHYSICAL CHEMISTRY DIVISION
Quantities, Units and Symbols in Physical Chemistry
Prepared for publication by
IAN MILLS
TOMISLAV CVITAŠ
KLAUS HOMANN
NIKOLA KALLAY
KOZO KUCHITSU
SECOND EDITION
BLACKWELL SCIENCE



Unit Conversion

7.2 CONVERSION TABLES FOR UNITS

The table below gives conversion factors from a variety of units to the corresponding SI unit. Examples of the use of this table have already been given in the preceding section. For each physical quantity the name is given, followed by the recommended symbol(s). Then the SI unit is given, followed by the esu, emu, Gaussian unit (Gau), atomic unit (au), and other units in common use, with their conversion factors to SI. The constant ζ which occurs in some of the electromagnetic conversion factors is the (exact) pure number $2.997\,924\,58 \times 10^{10} = c_0/(\mathrm{cm\ s^{-1}})$.

The inclusion of non-SI units in this table should not be taken to imply that their use is to be encouraged. With some exceptions, SI units are always to be preferred to non-SI units. However, since many of the units below are to be found in the scientific literature, it is convenient to tabulate their relation to the SI.

For convenience units in the esu and Gaussian systems are quoted in terms of the four dimensions *length*, *mass*, *time*, and *electric charge*, by including the franklin (Fr) as an abbreviation for the electrostatic unit of charge and $4\pi\varepsilon_0$ as a constant with dimensions $(charge)^2/(energy \times length)$. This gives each physical quantity the same dimensions in all systems, so that all conversion factors are pure numbers. The factors $4\pi\varepsilon_0$ and the Fr may be eliminated by writing Fr = esu of charge = $\mathrm{erg^{1/2}\,cm^{1/2}} = \mathrm{cm^{3/2}\,g^{1/2}\,s^{-1}}$, and $4\pi\varepsilon_0 = \zeta^2 = 1\ \mathrm{Fr^2\ erg^{-1}\,cm^{-1}} = 1$, to recover esu expressions in terms of three base units (see section 7.3 below). The symbol Fr should be regarded as a compact representation of (esu of charge).

Conversion factors are either given exactly (when the = sign is used), or they are given to the approximation that the corresponding physical constants are known (when the ≈ sign is used). In the latter case the uncertainty is always less than ± 5 in the last digit quoted.

Name	*Symbol*	*Relation to SI*
length, l		
metre (SI unit)	m	
centimetre (cgs unit)	cm	$= 10^{-2}$ m
bohr (au)	a_0, b	$= 4\pi\varepsilon_0\hbar^2/m_e e^2 \approx 5.291\,77 \times 10^{-11}$ m
ångström	Å	$= 10^{-10}$ m
micron	μ	$= \mu\mathrm{m} = 10^{-6}$ m
millimicron	mμ	$= \mathrm{nm} = 10^{-9}$ m
x unit	X	$\approx 1.002 \times 10^{-13}$ m
fermi	f, fm	$= \mathrm{fm} = 10^{-15}$ m
inch	in	$= 2.54 \times 10^{-2}$ m
foot	ft	$= 12$ in $= 0.3048$ m
yard	yd	$= 3$ ft $= 0.9144$ m
mile	mi	$= 1760$ yd $= 1609.344$ m
nautical mile		$= 1852$ m
astronomical unit	AU	$= 1.496\,00 \times 10^{11}$ m

Abbreviations

AA	atomic absorption
AAS	atomic absorption spectroscopy
ac	alternating current
ACM	adiabatic channel model
ACT	activated complex theory
A/D	analog-to-digital
ADC	analog-to-digital converter
AES	Auger electron spectroscopy
AIUPS	angle-integrated ultraviolet photoelectron spectroscopy
AM	amplitude modulated
amu	atomic mass unit (symbol: u) (see p.75)
AO	atomic orbital
APS	appearance potential spectroscopy
ARAES	angle-resolved Auger electron spectroscopy
AS	Auger spectroscopy
ATR	attenuated total (internal) reflection
AU	astronomical unit (see p. 110)
au	atomic unit (see section 7.3, p.120)
bcc	body centred cubic
BET	Brunauer–Emmett–Teller
BIS	bremsstrahlung isochromat spectroscopy
BM	Bohr magneton (symbol: μ_B, see p.116)
bp	boiling point
Btu	British thermal unit (see p.112)

Quantities

2.5 ATOMS AND MOLECULES

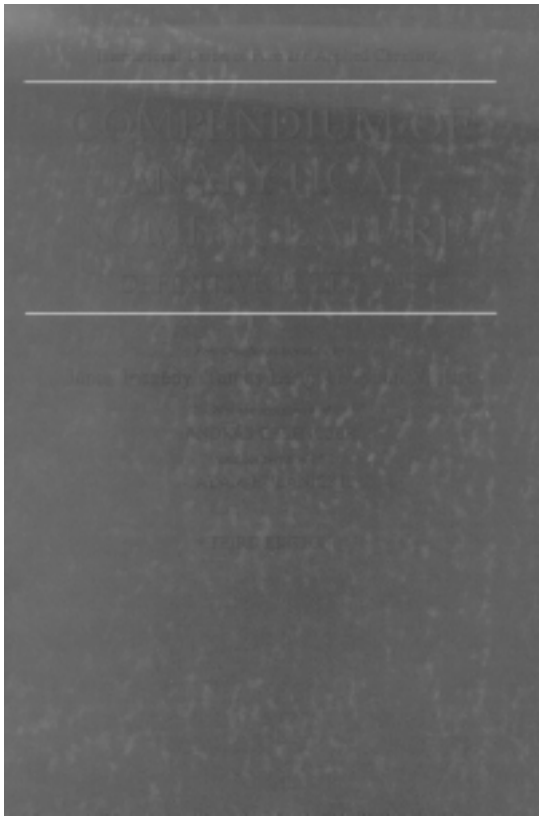
The names and symbols recommended here are in agreement with those recommended by IUPAP [4] and ISO [5.j]. Additional quantities and symbols used in atomic, nuclear and plasma physics can be found in [4 and 5.k].

Name	*Symbol*	*Definition*	*SI unit*	*Notes*
nucleon number, mass number	A		1	
proton number, atomic number	Z		1	
neutron number	N	$N = A - Z$	1	
electron rest mass	m_e		kg	1, 2
mass of atom, atomic mass	m_a, m		kg	
atomic mass constant	m_u	$m_u = m_a(^{12}C)/12$	kg	1, 3
mass excess	Δ	$\Delta = m_a - Am_u$	kg	
elementary charge, proton charge	e		C	2
Planck constant	h		J s	
Planck constant/2π	$\hbar$	$\hbar = h/2\pi$	J s	2
Bohr radius	a_0	$a_0 = 4\pi\varepsilon_0\hbar^2/m_e e^2$	m	2
Hartree energy	E_h	$E_h = \hbar^2/m_e a_0^2$	J	2
Rydberg constant	R_∞	$R_\infty = E_h/2hc$	m^{-1}	
fine structure constant	α	$\alpha = e^2/4\pi\varepsilon_0\hbar c$	1	
ionization energy	E_i		J	
electron affinity	E_{ea}		J	
electronegativity	χ	$\chi = \frac{1}{2}(E_i + E_{ea})$	J	4
dissociation energy	E_d, D		J	
from the ground state	D_0		J	5
from the potential minimum	D_e		J	5

(1) Analogous symbols are used for other particles with subscripts: p for proton, n for neutron, a for atom, N for nucleus, etc.
(2) This quantity is also used as an atomic unit; see sections 3.8 and 7.3.
(3) m_u is equal to the unified atomic mass unit, with symbol u, i.e. $m_u = 1$ u (see section 3.7). In biochemistry the name dalton, with symbol Da, is used for the unified atomic mass unit, although the name and symbol have not been accepted by CGPM.

Elements

Z	*Symbol*	*A*	*Atomic mass,* m_a/u	*Isotopic abundance,* 100 *x*	*Nuclear spin,* *I*	*Magnetic moment,* m/μ_N	*Quadrupole moment,* Q/fm^2
7	N	14	14.003 074 002 (26)	99.634 (9)	1	+0.403 761 00 (6)	+2.01 (2)
		15	15.000 108 97 (4)	0.366 (9)	1/2	−0.283 188 842 (45)	
8	O	16	15.994 914 63 (5)	99.762 (15)	0	0	
		17	16.999 1312 (4)	0.038 (3)	5/2	−1.893 80	−2.558 (22)
		18	17.999 1603 (9)	0.200 (12)	0	0	
9	F	19	18.998 403 22 (15)	100	1/2	+2.628 868 (8)	
10	Ne	20	19.992 4356 (22)	90.48 (3)	0	0	
		21	20.993 8428 (21)	0.27 (1)	3/2	−0.661 797 (5)	+10.155 (75)
		22	21.991 3831 (18)	9.25 (3)	0	0	
11	Na	23	22.989 7677 (10)	100	3/2	+2.217 6556 (6)*	+10.06 (20)
12	Mg	24	23.985 0423 (8)	78.99 (3)	0	0	
		25	24.985 8374 (8)	10.00 (1)	5/2	−0.855 465 (8)	+19.94 (20)
		26	25.982 5937 (8)	11.01 (2)	0	0	
13	Al	27	26.981 5386 (8)	100	5/2	+3.641 504 687 (65)	+14.03 (10)
14	Si	28	27.976 9271 (7)	92.23 (1)	0	0	
		29	28.976 4949 (7)	4.67 (1)	1/2	−0.555 29 (3)	
		30	29.973 7707 (7)	3.10 (1)	0	0	
15	P	31	30.973 7620 (6)	100	1/2	+1.131 60 (3)	
16	S	32	31.972 070 70 (25)	95.02 (9)	0	0	
		33	32.971 458 43 (23)	0.75 (1)	3/2	+0.643 8212 (14)	−6.78 (13)
		34	33.967 866 65 (22)	4.21 (8)	0	0	
		36	35.967 080 62 (27)	0.02 (1)	0	0	
17	Cl	35	34.968 852 721 (69)	75.77 (5)	3/2	+0.821 8743 (4)	−8.11 (8)
		37	36.965 902 62 (11)	24.23 (5)	3/2	+0.684 1236 (4)	−6.39 (6)
18	Ar	36	35.967 545 52 (29)	0.337 (3)	0	0	
		38	37.962 7325 (9)	0.063 (1)	0	0	
		40	39.962 3837 (14)	99.600 (3)	0	0	
19	K	39	38.963 7074 (12)	93.2581 (44)	3/2	+0.391 507 31(12)*	+5.9 (6)
		40	39.963 9992 (12)	0.0117 (1)	4	−1.298 1003 (34)	−7.3 (7)
		41	40.961 8254 (12)	6.7302 (44)	3/2	+0.214 870 09 (22)	+7.2 (7)



Logarithm of the Retention Factor

This term is equivalent to the R_M value used in planar chromatography (see *R_M value*). The symbol κ is suggested to express log k:

$$\kappa = \log k = \log [(1 - R)/R]$$

Retardation Factor (R)

The fraction of the sample component in the mobile phase at equilibrium; it is related to the retention factor and other fundamental chromatography terms:

$$R = 1/(k + 1)$$

Relative Retention (r)

The ratio of the adjusted or net retention volume (time) or retention factor of a component relative to that of a standard, obtained under identical conditions:

$$r = V_{Ri}'/V_{R(st)}' = V_{Ni}/V_{N(st)} = t_{Ri}'/t_{R(st)}' = k_i/k_{st}$$

Depending on the relative position of the peak corresponding to the standard compound in the chromatogram, the value of r may be smaller, larger or identical to unity.

TABLE 10.9 Transformation of sample into vapour.
Terms symbols and units for measurable quantities

Terms	Symbol	Practical Unit	Note
Rate of liquid consumption	F_l	cm^3s^{-1}	In the usual case of a pneumatic nebulizer F_l is called the rate of liquid aspiration
Efficiency of nebulization	ϵ_n	1	
Fraction desolvated	β_s	1	
Fraction volatilized	β_v	1	
Fraction atomized	β_a	1	
Efficiency of atomization	ϵ_a	1	
Flame temperature	T_F	K	When the temperature varies

12.3.3.2 Tandem Mass Spectrometers

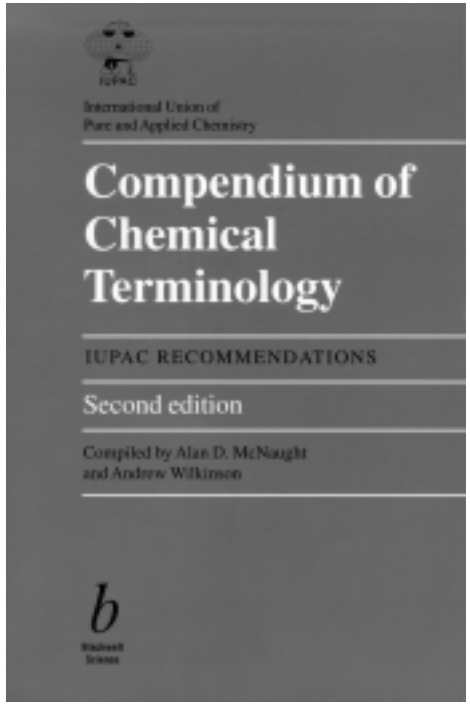
General

Fixed neutral loss (gain) scan

A scan that determines, in a single instrument, all the parent ion mass/charge ratios which react to the loss or gain of a selected neutral mass.

Fixed neutral loss (gain) spectrum

A spectrum obtained when data are acquired that determine all the parent ion mass/charge ratios that react by the loss (gain) of a selected neutral mass.



nuclear fusion reaction
A reaction between two light nuclei resulting in the production of a nuclear species heavier than either initial nucleus.
1982, *54*, 1543

nuclear graphite
A *polygranular graphite* material for use in nuclear reactor cores consisting of *graphitic carbon* of very high chemical purity. High purity is needed to avoid absorption of low-energy neutrons and the production of undesirable radioactive species.
Notes:
Apart from the absence of neutron-absorbing impurities, modern reactor graphites are also characterized by a high degree of *graphitization* and no preferred bulk orientation. Such properties increase the dimensional stability of the nuclear graphite at high temperatures and in a high flux of neutrons. The term nuclear graphite is often, but incorrectly, used for any *graphite material* in a nuclear reactor, even if it serves only for structural purposes.
1995, *67*, 498

nuclear isomers
Nuclides having the same *mass number* and *atomic number,* but occupying different nuclear energy states.
1982, *54*, 1545

nuclear level
One of the energy values at which a *nucleus* can exist for an appreciable time ($> 10^{-22}$ s).
1982, *54*, 1547

nuclear magneton
Electromagnetic fundamental physical constant $\mu_N = (m_e/m_p)\mu_B = 5.050\ 7866\ (17) \times 10^{-27}$ J T^{-1}, where m_e is the electron rest mass, m_p the proton rest mass and μ_B the Bohr magneton.
CODATA Bull., 1986, *63*, 1

nuclear particle

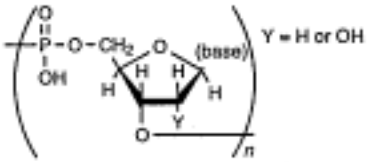
nuclear transition
For a *nucleus* a change from one quantized energy state into another or a *nuclear transformation.*
1982, *54*, 1553

nucleating agent
A material either added to or present in a system, which induces either homogeneous or heterogeneous *nucleation.*
1972, *31*, 608

nucleation (in colloid chemistry)
The process by which nuclei are formed in solution. The condensation of a single chemical compound is called homogeneous nucleation. The simultaneous condensation of more than one compound is called simultaneous nucleation. The condensation of a compound on a foreign substance is called heterogeneous nucleation.
O.B. 84; see also 1972, *31*, 608

nucleation and growth
A process in a *phase transition* in which nuclei of a new phase are first formed, followed by the propagation of the new phase at a faster rate.
See *continuous precipitation, discontinuous precipitation.*
1994, *66*, 587

nucleic acids
Macromolecules, the major organic matter of the nuclei of biological cells, made up of *nucleotide* units, and hydrolysable into certain *pyrimidine* or *purine bases* (usually adenine, cytosine, guanine, thymine, uracil), D-ribose or 2-deoxy-D-ribose and phosphoric acid.



Varieties of Terms in IUPAC Glossaries

- Data
  - Numeric Quantities (number + dimension)
  - Non-numeric
    - Formulas, reactions, phases, structure
- Concepts
  - Chemicals Properties and Behavior
    - Chemical rules, properties, attributes and behavior
  - Methods
    - Apparatus, procedures, acronyms
  - Reporting
    - Process and Represent Results
  - Non-chemical concepts
    - Use other namespaces one day

What to do?

- Publish glossaries as XML instances (documents)
- Restructure glossaries to serve as ‘namespace’
  - Utilize CML-STM
- Express implicit constraints for data checking
  - Enforce quantity/dimension relations
- (Re)organize data structure
  - Fit neatly in a tree

Publish Glossaries in XML

- Benefits
  - Forces a logical organization on the information
  - Lead by example
  - Supply test documents to developers
  - Information first, uses follow
- Negatives
  - Little present ‘pull’ from users
    - Is it too early?
    - Paucity of end user software tools
  - Leaves too much work

Restructure Glossaries

- Very basic ‘namespace’ required for further growth
- Establish IUPAC as source of XML chemical terminology.

- Is it too early?

namespace

applications

XML ‘Instance’

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<glossary xmlns=http://www.iupac.org/GreenBook
xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
xsi:schemaLocation="http://www.iupac.org/GreenBook D:\test.xsd">

<quantity name="cartesianCoordinates" pname="Cartesian Space Coordinates" dictRef=""
class="spaceAndTime">

        <i:dimension system="si">
                <i:unit name="m"/>
        </i:dimension>

        <i:symbolSet class="any">
                <i:symbol>x</i:symbol>
                <i:symbol>y</i:symbol>
                <i:symbol>z</i:symbol>
        </i:symbolSet>

</quantity>
```

Constraints for Data Checking

- Provide XML Schema

- Physical Quantities
  - Ensure correct dimensions
- Chemical Quantities
  - Valid formulas, state symbols,
  - Chemical Constants
    - Periodic Table

```
<schema targetNamespace="http://www.iupac.org/GreenBook"
xmlns:i="http://www.iupac.org/GreenBook"
xmlns="http://www.w3.org/2001/XMLSchema"
elementFormDefault="unqualified"
attributeFormDefault="unqualified">

<annotation>
        <documentation>

IUPAC test Schema

        </documentation>
</annotation>

<!-- examples of prefix/power templates -->

<complexType name="dimensionPrePowType">
        <attribute name="prefix" type="string" use="optional"/>
        <attribute name="pow" type="string" use="required"/>
        <!-- restrict these later to valid values -->
</complexType>


<complexType name="pow1Type">
        <complexContent>
                <restriction base="i:dimensionPrePowType">
                        <attribute name="pow" use="required" fixed="1"/>
                </restriction>
        </complexContent>
</complexType>
```

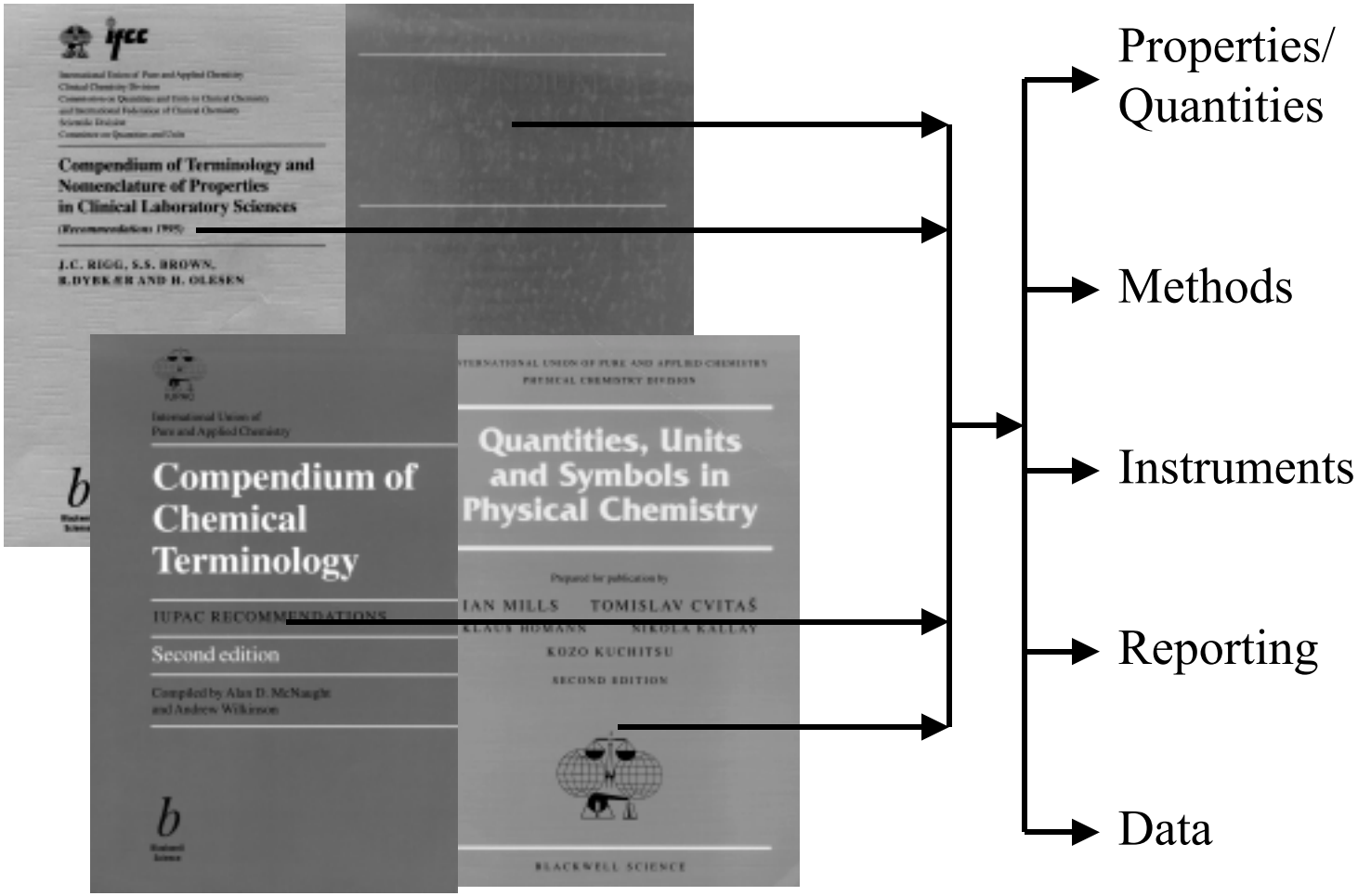
```
<?xml version="1.0" encoding="ISO-8859-1"?>

<physicalProperty xmlns="http://www.iupac.org/GreenBook"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.iupac.org/GreenBook D:\test.xsd">

<i:energy type="pure">
        <i:value>
                <i:nvalue>12.2</i:nvalue>
                <i:uncert>
                        <i:nsignif>3</i:nsignif>
                </i:uncert>
        </i:value>
        <i:units>
                <i:joule pow="1"/>
        </i:units>
</i:energy>


</physicalProperty>
```

(Re)organize Data Structure?



Help Wanted

- ‘Domain’ Experts
- Data Distributors
- XML/Software Developers

- ‘International’UPAC

- steve.stein@nist.gov