

3. CIF DATA DEFINITION AND CLASSIFICATION

Table 3.1.7.1. Abbreviations in CIF data names

Terms for which abbreviations are defined are sometimes found unabbreviated.

Abbreviation	Term	Abbreviation	Term	Abbreviation	Term
abbrev	abbreviation	eqn	equation	oper	operation
abs	absolute (configuration, not structure)	esd	standard uncertainty (estimated standard deviation) (<i>see su</i>)	org	organism
absorpt	absorption			orient	orientation
alt	alternative	expt	experiment	origx	orthogonal coordinate matrix (PDB files)
amp	amplitude	exptl	experimental	os	operating system
AN	accession number	fom	figure of merit	param	parameter
anal	analyser	fract	fractional	pd	powder diffraction
aniso	anisotropic*	Fsqd	<i>F</i> squared	PDB	Protein Data Bank
anisotrop	anisotropic*	gen	generation	PDF	Powder Diffraction File
anom	anomalous	gen	generator	perp	perpendicular
ASTM	American Society for Testing and Materials	gen	genetic	phos	phosphate
asym	asymmetric	geom	geometric	pk	peak
atten	attenuation	H-M	Hermann–Mauguin	polarisn	polarization
au	arbitrary units	ha	heavy atom	poly	polymer
auth	author	hbond	hydrogen bond	pos	position
av	average	hist	history	prep	preparation
ax	axial	horiz	horizontal	proc	processed
B	<i>B</i> form of atomic displacement parameter (a.d.p.)	I	intensity	prof	profile
backgd	background*	ICSD	Inorganic Crystal Structure Database	prot	protein
beg	begin	id	identifier	ptnr	partner
bg	background*	illum	illumination	publ	publication
biol	biology	imag	imaginary	R	agreement index
bkg	background*	inc	increment	rad	radius
bond	bonding	incl	include	recd	received
Bsol	<i>B</i> form of a.d.p. for solvent	info	information	recip	reciprocal
calc	calculated	instr	instrument	ref	reference
calib	calibration (pd)	Int	international	refine	refinement
cartn	Cartesian	ISBN	International Standard Book Number	refln	reflection
CAS	<i>Chemical Abstracts Service</i>	iso	isotropic	reflns	reflections
char	characterization (pd)	iso	isomorphous	res	resolution
chem	chemical	ISSN	International Standard Serial Number	restr	restraints
chir	chirality	IUCr	International Union of Crystallography	rev	revision
clust	cluster	IUPAC	International Union of Pure and Applied Chemistry	Rmerge	agreement index of merging
coef	coefficient	len	length	rms	root mean square
com	common	lim	limit	rot	rotation
comp	component	loc	lack of closure	S	goodness of fit
conc	concentration	ls	least squares	samp	sample
conf	conformation	max	maximum	scat	scattering factor
config	configuration	MDF	Metals Data File	seq	sequence
conform	conformant	meanI	mean intensity	sigI	$\sigma(I)^*$
conn	connectivity	meas	measured	sigmaI	$\sigma(I)^*$
cons	constant	mid	middle (between max and min)	sint	$\sin \theta$
CSD	Cambridge Structural Database	min	minimum	sint/lambda	$\sin(\theta)/\lambda^*$
db	database	mod	modification	sol	solvent
defn	definition	mods	modifications	spec	specimen
detc	detector	mon	monomer	src	source
der	derivative	monochr	monochromator (pd)*	std	standard
dev	standard deviation	mono	monochromator (pd)*	stol	$\sin(\theta)/\lambda^*$
dict	dictionary	nat	natural	struct	structure
dif	difference*	NBS	National Bureau of Standards (now National Institute of Standards and Technology)	su	standard uncertainty
diff	difference*			suppl	supplementary
diffr	diffractometer			sys	systematic
diffm	diffraction	NCA	number of connected atoms	tbar	mean path length
displace	displacement	ncs	noncrystallographic symmetry	temp	temperature
dist	distance	netI	net intensity	tor	torsion angle
divg	divergence	NH	number of connected hydrogen atoms	tran	transformation*
dom	domain	nha	non-hydrogen atoms	transf	transformation*
dtime	deadtime	norm	normal	transform	transformation*
ens	ensemble	nst	nonstandard	tvect	translation vector (PDB files)
eq	equatorial*	nucl	nucleic acid	vert	vertical
equat	equatorial*	num	number	wR	weighted agreement index
equiv	equivalent	obs	observed	wt	weight

* Terms with multiple definitions.

It is clear that the location specified for the local dictionary is only meaningful for applications running on the same computer or network, and therefore the ability to validate against this local dictionary is not portable. On the other hand, it may be that the local data names used by the authors of this CIF are not intended to have meaning outside their own laboratory.

3.1.8.2. The dictionary register

COMCIFS maintains a register of dictionaries known to it, including the identifying name and version strings within those dictionaries. The register also includes the location of each dictionary, expressed at present as a URL designed to allow retrieval

by file transfer protocol (ftp) from the IUCr server. Changes in the location of a particular dictionary file can be made by modifying the entry in the register, avoiding the problem of specifying a URL in a data file that would then become outdated if the dictionary was moved. Dictionary applications can consult the register (according to a protocol outlined below) to locate and retrieve the dictionaries needed for validating data files. It is of course essential that the validation software knows how to locate the register. The location is at present given by the URL <ftp://ftp.iucr.org/pub/cifdics/cifdic.register>.

The problem of changing URLs has therefore not disappeared completely, but is at least confined to the need to maintain one single address.

3.1. GENERAL CONSIDERATIONS WHEN DEFINING A CIF DATA ITEM

Table 3.1.8.1. CIF dictionary register (maintained as a STAR File)

```

data_validation_dictionaries
loop_
  _cifdic_dictionary.name
  _cifdic_dictionary.version
  _cifdic_dictionary.DDL_compliance
  _cifdic_dictionary.reserved_prefix
  _cifdic_dictionary.URL
  _cifdic_dictionary.description
cif_core.dic . 1.4.1 .
  ftp://ftp.iucr.org/pub/cifdics/cif_core.dic
  'Core CIF Dictionary'
cif_core.dic 1.0 . .
  ftp://ftp.iucr.org/pub/cifdics/cifdic.C91
  'Original Core CIF Dictionary'
cif_core.dic 2.3.1 1.4.1 .
  ftp://ftp.iucr.org/pub/cifdics/cif_core_2.3.1.dic
  'Core CIF Dictionary'
cif_pd.dic . 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_pd.dic
  'Powder CIF Dictionary'
cif_pd.dic 1.0.1 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_pd_1.0.1.dic
  'Powder CIF Dictionary'
cif_ms.dic . 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_ms.dic
  'Modulated structures CIF Dictionary'
cif_ms.dic 1.0.1 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_ms_1.0.1.dic
  'Modulated structures CIF Dictionary'
cif_rho.dic . 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_rho.dic
  'Modulated structures CIF Dictionary'
cif_rho.dic 1.0.1 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_rho_1.0.1.dic
  'Electron density CIF Dictionary'
cif_mm.dic . 2.1.2 .
  ftp://ftp.iucr.org/pub/cifdics/cif_mm.dic
  'Macromolecular CIF Dictionary'
cif_mm.dic 1.0 2.1.2 .
  ftp://ftp.iucr.org/pub/cifdics/cif_mm_1.0.dic
  'Macromolecular CIF Dictionary'
mmcif_std.dic . 2.1.6 .
  ftp://ftp.iucr.org/pub/cifdics/mmcif_std.dic
  'Macromolecular CIF Dictionary'
mmcif_std.dic 2.0.09 2.1.6 .
  ftp://ftp.iucr.org/pub/cifdics/cif_mm_2.0.09.dic
  'Macromolecular CIF Dictionary'
cif_img.dic . 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/cif_img.dic
  'Image CIF Dictionary'
cif_img.dic 1.0 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/cif_img_1.0.dic
  'Image CIF Dictionary'
cif_img.dic 1.3.2 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/cif_img_1.3.2.dic
  'Image CIF Dictionary'
cif_sym.dic . 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/cif_sym.dic
  'Symmetry CIF Dictionary'
cif_sym.dic 1.0.1 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/cif_sym_1.0.1.dic
  'Symmetry CIF Dictionary'
cif_compat.dic . 1.4 .
  ftp://ftp.iucr.org/pub/cifdics/cif_compat.dic
  'Legacy CIF Dictionary of deprecated terms'
ddl_core.dic . 1.4.1 .
  ftp://ftp.iucr.org/pub/cifdics/ddl_core.dic
  'Non-relational dictionary definition language'
ddl_core_2.1.3.dic . 2.1.3 .
  ftp://ftp.iucr.org/pub/cifdics/ddl_core_2.1.3.dic
  'Relational dictionary definition language'
mmcif_ddl.dic . 2.1.6 .
  ftp://ftp.iucr.org/pub/cifdics/mmcif_ddl.dic
  'Relational dictionary definition language'

```

Table 3.1.8.1 is an extract from the current register. The latest version of the register will always be available from the URL given above.

The entries for each dictionary include one with the version string set to '.', representing the current version; this is the version that should be retrieved unless a data file specifies otherwise.

Note that the register may also contain locators for local dictionaries constructed by owners of reserved prefixes (Section 3.1.2.2) when the owner has requested that a dictionary of local names be made publicly available. An appropriate name for a local dictionary in the register (`_dictionary_name` or `_dictionary.title` for DDL1 or DDL2 dictionaries, respectively) would be `cif_local_myprefix.dic`, where the string indicated by `myprefix` is one of the prefixes reserved for private use by the author of the dictionary (see Section 3.1.2.2). This scheme complements the naming convention for public dictionaries.

3.1.8.3. Locating a dictionary for validation

The following protocol applies to the creation and use of software designed to locate the dictionaries referenced by a data file and validate the data file against them. The protocol is necessary to address the issues that arise because dictionaries evolve through various audited versions, because not all dictionaries referenced by a data file may be accessible, and because data files might not in practice contain pointers to their associated dictionaries.

Software source code for applications that use CIF dictionaries to validate the contents of data files should be distributed with a copy of the most recent version of the register of dictionaries, and with the URL of the master copy hard-coded. Library utilities should be provided that permit local cacheing of the register file and the ability to download and replace the cached register at regular intervals. Individual dictionary files located and retrieved through the use of the register should also be cached locally, to guard against temporary unavailability of network resources.

Each CIF data file should contain a reference to one or more dictionary files against which the file may be validated. At the very least this will be `_audit_conform_dict_name` (`_audit_conform.dict_name` for DDL2 files) (N). `*_version` (V) and `*_location` (L) are optional. In the event that no dictionaries are specified, the default validation dictionary should be that identified as having $N = \text{cif_core.dic}$ and $V = \text{'.'}$ (i.e. the most recent version of the core dictionary). Since dictionaries are intended always to be extended, it is normally enough just to specify the name (and possibly the location).

This default is appropriate for most well formed CIFs, but if it is important to provide formal validation of old CIFs conforming to the earliest printed specification, which used the now-deprecated units extension convention, the dictionary `cif.compat.dic` may also be added to the default list (Section 3.1.5.4.3).

There is a difficulty associated with assuming this default for CIFs containing DDL2 data names. At present, the DDL2 version of the core dictionary does not exist as a separate file. Most existing CIFs built on the DDL2 model conform to the macromolecular (mmCIF) dictionary, and so best current working practice is to assume a default validation dictionary for DDL2-style CIFs with $N = \text{mmcif_std.dic}$ and $V = \text{'.'}$ (i.e. the most recent version of the mmCIF dictionary), since this includes the core data names as a subset. However, to anticipate future developments, it is suggested that applications built to validate DDL2 files first search the register for a default entry with $N = \text{cif_core.dic}$, $V = \text{'.'}$ and a value of 2 or higher for the relevant DDL version:

```

loop_
  _cifdic_dictionary.name
  _cifdic_dictionary.version
  _cifdic_dictionary.DDL_compliance
cif_core.dic . 2.1.2

```