

# Dictionary of XAS Data Interchange Metadata

## XDI Working Group

- Matthew NEWVILLE (University of Chicago Center for Advanced Radiation Sources, APS)
- Bruce RAVEL (NIST) [bravel AT bnl DOT gov](mailto:bravel@bnl.gov)
- V. Armando SOLÉ (ESRF)
- Gerd WELLENREUTHER (DESY)
- mailing list - <http://millenia.cars.aps.anl.gov/mailman/listinfo/xasformat>
- GitHub organization - <https://github.com/XraySpectroscopy>

## Contents

<b>1</b>	<b>Overview</b>	<b>2</b>
1.1	The meaning of metadata . . . . .	2
1.2	The <i>XDI</i> syntax . . . . .	2
1.3	The format of the value . . . . .	2
<b>2</b>	<b>The dictionary</b>	<b>4</b>
2.1	Name spaces . . . . .	4
2.2	Tags . . . . .	4
2.3	Required metadata . . . . .	4
2.4	Recommeneded metadata . . . . .	5
2.5	Defined items in the Facility namespace . . . . .	5
2.6	Defined items in the Beamline namespace . . . . .	6
2.7	Defined items in the Mono namespace . . . . .	6
2.8	Defined items in the Detector namespace . . . . .	6
2.9	Defined items in the Sample namespace . . . . .	7
2.10	Defined items in the Scan namespace . . . . .	8
2.11	Defined items in the Element namespace . . . . .	8
2.12	Defined items in the Column namespace . . . . .	9
2.13	Extension fields . . . . .	11

## 1 Overview

This is **version 1.0** of the dictionary of metadata to be used with the XAS Data Interchange (XDI) format. Each item definition includes:

1. The name representing the datum
2. The meaning of the datum
3. Thw units of the datum
4. The format for representing its value

Words used to signify the requirements in the specification **shall** follow the practice of [RFC 2119](#).

A use of this dictionary is not compliant if it fails to satisfy one or more of the **must** or **required** level requirements presented herein.

### 1.1 The meaning of metadata

The purpose of this dictionary is to identify a set of metadata to be encoded in the specification of the *XDI* format and to assign names to each meaningful concept. This effort must take a broad view, capturing metadata concepts as broadly as they are used in the community. This effort must also be open ended in that there must be a mechanism for providing new forms of metadata not considered up front. This effort is intended to serve as the XAS metadata dictionary for other data format types, for instance a database format for libraries of XAS spectra or a hierarchical format for multi-spectral datasets.

### 1.2 The *XDI* syntax

This dictionary has been developed along with the [XDI specification](#). All examples given in this dictionary use all recommendations of the *XDI* syntax. The metadata name consists of the capitalized namespace, followed by a dot, followed by a tag. Here is an example: `Element.symbol`. When appearing in an *XDI* file to convey a metadata value, the line begins with a comment token and end with an end-of-line token. A colon is the delimiting token between the metadata name and its value. Here is an example:

```
# Element.symbol: Cu
```

### 1.3 The format of the value

Some of the tags in this dictionary have formatted values as part of their definitions.

- *string*: A string is specifically an ASCII string represented by printable characters in the lower 128 of the ASCII set. This **must** be the English-language representation of the value. For example, the string representing `Facility.name` for the Thai synchrotron **must** be `SLRI` rather than a sequence of characters in the Thai script.

- *free-format string*: This is a string which can contain any character (save end-of-line characters) in any encoding system. A free-format string need not be ASCII and need not be English. Because applications using *XDI* may not be capable of handling some encoding systems, it is **recommended** that free-format strings be ASCII.
- *string + units*: This is a string as defined above, followed by white space, followed by a string denoting the units of the previous string. As an example, a value for `Column.1` might be `energy eV`, which identifies the contents of the first column in the data table as containing energy values expressed in electron volt units. The selection of possible units for a tag is given in the definition of the tag.
- *float*: A float is a string which is interpretable as a floating-point number in the C programming language. An integer is permissible. Values of `NaN`, `sNaN`, `qNaN`, `inf`, `+inf`, and `-inf` are not allowed in *XDI*. That is, a float in *XDI* **must** be a finite number. See [IEEE 754-2008](#).
- *float + units*: This is a float as defined above, followed by white space, followed by a string identifying the units of the number. For example, a value for `Sample.temperature`, which identifies the temperature at which an XAS measurement is made, might be `500 K`, identifying the temperature of the measurements in Kelvin temperature units. The selection of possible units for a tag is given in the definition of the tag.
- *chemical formulas*: `Sample.stoichiometry` is intended to represent the elemental composition of the sample. To allow interpretation of chemical formulas by computer, this field and extension fields which represent chemical information **must** use the [IUCr definition of a chemical formula](#).
- *time*: Because of the wide variability of cultural standards in the representation of time, *XDI* defines a strict standard for time stamps in *XDI* files. `Scan.start_time`, `Scan.end_time`, and any extension fields dealing in time **must** use the [ISO 8601 specification for combined dates and times](#)
- *element symbols*: `Element.symbol`, `Element.reference`, and any extension fields identifying specific elements **must** use one of the recognized 1, 2, or 3 letter symbols given in Sec. 2.11
- *edge symbols*: `Element.edge`, `Element.ref_edge`, and any extension fields identifying specific absorption edges **must** use one of the recognized 1 or 2 letter symbols given in Sec. 2.11. Note that the subscript is represented as an Arabic numeral and not as a Roman numeral.

Some additional comments:

- Locale is **not** respected when interpreting floating point numbers. The decimal mark **must** be a dot (.<sup>1</sup>, ASCII 46). The decimal mark **must not** be a comma (,<sup>1</sup>, ASCII 44).
- A tag which is in a defined family but which is not defined in this dictionary **must** be interpreted as have a free-format string as its value.
- A tag which is present in an *XDI* file but which has no value or only white space as its value (i.e. the colon is followed by zero or more spaces tokens then by an end-of-line token) **must** be interpreted as a zero-length string or as the value 0, as appropriate to the value type.
- Strings identifying facilities and beamlines **must** use whatever convention is in use at the beamline. In the case where a beamline is known both by a designation and a name (for example, beamline 13ID at the Advanced Photon Source is also known by its name, "GSECARS"), the designation is **recommended**.

## 76      2 The dictionary

### 77      2.1 Name spaces

78      The purpose of namespaces is to provide sensible, widely understood, semantic groupings of defined  
 79      metadata tags. All tags associated with conveying information about sample preparation and the  
 80      measurement environment of the sample belong in the *Sample* namespace, all tags associated with  
 81      the configuration of the beamline optics belong in the *Beamline* namespace, and so on.

82      Namespaces are strings composed of a subset of the ASCII character set. The first character **must**  
 83      be a letter. The remaining characters **must** be letters, numbers, underscores, or dashes. Letters are  
 84      ASCII 65 through 90 (`A-Z`) and ASCII 97-122 (`a-z`). Numbers are ASCII 48-57 (`0-9`). Underscore  
 85      (`_`) is ASCII 95 and dash (`-`) is ASCII 45. The namespace **must** be interpreted as case insensitive.

86      Here is a list of all defined semantic groupings:

- 87      1. `Facility`: Tags related to the synchrotron or other facility at which the measurement was  
       made
- 88      2. `Beamline`: Tags related to the structure of the beamline and its photon delivery system
- 89      3. `Mono`: Tags related to the monochromator
- 90      4. `Detector`: Tags related to the details of the photon detection system
- 91      5. `Sample`: Tags related to the details of sample preparation and measurement
- 92      6. `Scan`: Tags related to the parameters of the scan
- 93      7. `Element`: Tags related to the absorbing atom
- 94      8. `Column`: Tags used for identifying the data columns and their units

96      Below, specific members of these namespaces are defined. The definitions are not exclusive. Other  
 97      metadata can be placed in these namespaces as needed. Of course, undefined metadata are un-  
 98      likely to be interpreted correctly by applications using this dictionary. Metadata added to a defined  
 99      namespace **must not** use a defined tag. The defined namespaces and tags **shall** be interpreted  
 100     without sensitivity to case.

101     When defined metadata are present, the units and formatting specified below **must** be observed.

### 102     2.2 Tags

103     Tags are the words used to denote a specific entry in a namespace.

104     Tags are strings composed of a subset of the ASCII character set. All characters **must** be letters  
 105     (ASCII 65 through 90, `A-Z`) and ASCII 97-122, (`a-z`), numbers (ASCII 48-57, `0-9`), underscore  
 106     (ASCII 95, `_`), or dash (ASCII 45, `-`).

107     The tag **must** be interpreted as case insensitive.

### 108     2.3 Required metadata

109     Three items are essential to the interchange and successful interpretation of XAS data. These are  
 110     **required** for a file to be a compliant *XDI* file.

- 111     • `Element.symbol`: The element of the absorbing atom. The periodic table is replete with exam-  
       ples of atoms that have absorption edges with very similar edge energies. For example, the  
       tabulated values of the Cr K edge and the Ba L1 edge are both 5989 eV, while Se K and Tl  
       L3 are both at 12658. Without identification of the species of the absorbing atom and of the  
       absorption edge measured, some data cannot be unambiguously identified.

- 116     • `Element.edge`: The absorption edge measured. See above.
- 117     • `Mono.d_spacing`: The d-spacing of the monochromator. It is required to convert an abscissa  
118       represented as monochromator angle or encoder step count into energy. Also a correction  
119       to the energy axis of measured data, which may be required in the case of a miscalibration  
120       due to inaccuracies in the translation from angular position of the monochromator to energy,  
121       would need the d-spacing.

122     Most other metadata definitions that follow are **optional** for use with *XDI*. Some are **recommended**  
123     for use with all *XDI* files. The **recommended** metadata convey information that is of substantive  
124     value to the interpretation of the data.

## 125     2.4 Recommended metadata

126     The current list of recommended metadata, i.e. metadata which constitutes best practice when  
127     writing any data file, is

- 128     • `Facility.name`
- 129     • `Facility.xray_source`
- 130     • `Beamline.name`
- 131     • `Scan.start_time`
- 132     • `Column.1`

## 133     2.5 Defined items in the Facility namespace

- 134     • **Namespace:** `Facility` – **Tag:** `name`
  - 135       – *Description*: The name of synchrotron or other X-ray facility. This is **recommended** for  
136       use in all *XDI* files.
  - 137       – *Units*: none
  - 138       – *Format*: string
- 139     • **Namespace:** `Facility` – **Tag:** `energy`
  - 140       – *Description*: The energy of the stored current in the storage ring.
  - 141       – *Units*: GeV, MeV
  - 142       – *Format*: float + units
- 143     • **Namespace:** `Facility` – **Tag:** `current`
  - 144       – *Description*: The amount of stored current in the storage ring at the beginning of the  
145       scan.
  - 146       – *Units*: mA, A
  - 147       – *Format*: float + units
- 148     • **Namespace:** `Facility` – **Tag:** `xray_source`
  - 149       – *Description*: A string identifying the source of the X-rays, such as “bend magnet”, “undu-  
150       lator”, or “rotating copper anode”. This is **recommended** for use in all *XDI* files.
  - 151       – *Units*: none
  - 152       – *Format*: string

---

## 153 2.6 Defined items in the Beamline namespace

- 154 • Namespace: Beamline – Tag: name

155 – Description: The name by which the beamline is known. This is **recommended** for use  
 156 in all *XDI* files. For a beamline with a facility designation and a common name (such as  
 157 13-BM-B at the APS, also known as GSECARS), the designation is preferred.  
 158 – Units: none  
 159 – Format: free-format string

- 160 • Namespace: Beamline – Tag: collimation

161 – Description: A concise statement of how beam collimation is provided  
 162 – Units: none  
 163 – Format: free-format string

- 164 • Namespace: Beamline – Tag: focusing

165 – Description: A concise statement about how beam focusing is provided  
 166 – Units: none  
 167 – Format: free-format string

- 168 • Namespace: Beamline – Tag: harmonic\_rejection

169 – Description: A concise statement about how harmonic rejection is accomplished  
 170 – Units: none  
 171 – Format: free-format string

## 172 2.7 Defined items in the Mono namespace

- 173 • Namespace: Mono – Tag: name

174 – Description: A string identifying the material and diffracting plane or grating spacing of  
 175 the monochromator  
 176 – Units: none  
 177 – Format: free-format string

- 178 • Namespace: Mono – Tag: d\_spacing

179 – Description: The known d-spacing of the monochromator under operating conditions.  
 180 This is a **required** parameter for use with *XDI* when data are specified as a function of  
 181 angle or step count.  
 182 – Units: Å  
 183 – Format: float

184 This is the appropriate namespace for parameters of an energy dispersive polychromator. Such  
 185 parameters may be defined in future versions of this dictionary.

## 186 2.8 Defined items in the Detector namespace

- 187 • Namespace: Detector – Tag: i0

188 – Description: A description of how the incident flux was measured  
 189 – Units: none  
 190 – Format: free-format string

- 191 • Namespace: Detector – Tag: it

- 
- 192            - *Description*: A description of how the transmission flux was measured  
 193            - *Units*: none  
 194            - *Format*: free-format string
- 195        • **Namespace:** `Detector` – **Tag:** `if`  
           - *Description*: A description of how the fluorescence flux was measured  
 196            - *Units*: none  
 197            - *Format*: free-format string
- 198        • **Namespace:** `Detector` – **Tag:** `ir`  
           - *Description*: A description of how the reference flux was measured  
 199            - *Units*: none  
 200            - *Format*: free-format string

## 203        2.9 Defined items in the Sample namespace

- 204        • **Namespace:** `Sample` – **Tag:** `name`  
           - *Description*: A string identifying the measured sample  
 205            - *Units*: none  
 206            - *Format*: free-format string
- 207        • **Namespace:** `Sample` – **Tag:** `id`  
           - *Description*: A number or string uniquely identifying the measured sample. This is intended for interoperation with a database or laboratory management software. It could be, for example, a bar code number.  
 208            - *Units*: none  
 209            - *Format*: free-format string
- 210        • **Namespace:** `Sample` – **Tag:** `stoichiometry`  
           - *Description*: The stoichiometric formula of the measured sample  
 211            - *Units*: none  
 212            - *Format*: see the [IUCr definition of chemical\\_formula](#)
- 213        • **Namespace:** `Sample` – **Tag:** `prep`  
           - *Description*: A string summarizing the method of sample preparation  
 214            - *Units*: none  
 215            - *Format*: free-format string
- 216        • **Namespace:** `Sample` – **Tag:** `experimenters`  
           - *Description*: The names of the experimenters present for the measurement  
 217            - *Units*: none  
 218            - *Format*: free-format string
- 219        • **Namespace:** `Sample` – **Tag:** `temperature`  
           - *Description*: The temperature at which the sample was measured  
 220            - *Units*: degrees K, degrees C  
 221            - *Format*: float + units

222        The Sample namespace is rather open-ended. It is probably impossible to anticipate all the kinds of sample-related metadata that may be useful to attach to data. That said, it would be useful to suggest tags for a number of common kinds of extrinsic parameters along the line of `Sample.temperature`. These may be added as defined fields in future versions of the XDI specification.

- 
- 234     • `Sample.pressure`
  - 235     • `Sample.ph`
  - 236     • `Sample.eh`
  - 237     • `Sample.volume`
  - 238     • `Sample.porosity`
  - 239     • `Sample.density`
  - 240     • `Sample.concentration`
  - 241     • `Sample.resistivity`
  - 242     • `Sample.viscosity`
  - 243     • `Sample.electric_field`
  - 244     • `Sample.magnetic_field`
  - 245     • `Sample.magnetic_moment`
  - 246     • `Sample.crystal_structure`
  - 247     • `Sample.opacity`
  - 248     • `Sample.electrochemical_potential`

249     Many of these examples would take a float+units as values.

## 250     2.10 Defined items in the Scan namespace

- 251     • **Namespace:** `Scan` – **Tag:** `start_time`
  - 252       – *Description:* The beginning time of the scan. This is **recommended** for use with *XDI*.
  - 253       – *Units:* time
  - 254       – *Format:* [ISO 8601 specification for combined dates and times](#)
- 255     • **Namespace:** `Scan` – **Tag:** `end_time`
  - 256       – *Description:* The ending time of the scan.
  - 257       – *Units:* time
  - 258       – *Format:* [ISO 8601 specification for combined dates and times](#)
- 259     • **Namespace:** `Scan` – **Tag:** `edge_energy`
  - 260       – *Description:* The absorption edge as used in the data acquisition software.
  - 261       – *Units:* eV (**recommended**), keV, inverse Å
  - 262       – *Format:* float + units

263     This is the appropriate namespace for any parameters associated with scan parameters, such as  
264     integration times, monochromator speed, scan boundaries, or step sizes.

265     An example of a combined date and time representation is `2007-04-05T14:30:22`, which means 22  
266     seconds after 2:30 in the afternoon on the day of April 5th in the year 2007.

## 267     2.11 Defined items in the Element namespace

- 268     • **Namespace:** `Element` – **Tag:** `symbol`
  - 269       – *Description:* The measured absorption edge. This is a **required** parameter for use with  
270       *XDI*.
  - 271       – *Units:* none

- 272     – *Format*: one of these 118 1, 2, or 3 character strings for the standard atomic symbols  
 273     (not case sensitive):

```

274   H He Li Be B C N O F Ne Na Mg Al Si P S
275   Cl Ar K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge
276   As Se Br Kr Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd
277   In Sn Sb Te I Xe Cs Ba La Ce Pr Nd Pm Sm Eu Gd
278   Tb Dy Ho Er Tm Yb Lu Hf Ta W Re Os Ir Pt Au Hg
279   Tl Pb Bi Po At Rn Fr Ra Ac Th Pa U Np Pu Am Cm
280   Bk Cf Es Fm Md No Lr Rf Db Sg Bh Hs Mt Ds Rg Cn
281   Ut Fl Uup Lv Uus Uuo
  
```

282     See [Wikipedia's list of element symbols](#).

283 • **Namespace:** `Element` – **Tag:** `edge`

- 284     – *Description*: The measured absorption edge. This is a **required** parameter for use with  
 285     *XDI*.  
 286     – *Units*: none  
 287     – *Format*: one of these 28 1 or 2 character strings (not case sensitive):

```

288   K L L1 L2 L3 M M1 M2 M3 M4 M5
289   N N1 N2 N3 N4 N5 N6 N7 O 01 02 03 04 05 06 07
  
```

290     See table 10.10 at [IUPAC notation for X-ray absorption edges](#) for further explanation. The  
 291     use of the generic edges *L*, *M*, *N*, and *O* is **not recommended**, but **may** be used for spectra  
 292     spanning multiple edges.

293 • **Namespace:** `Element` – **Tag:** `reference`

- 294     – *Description*: The absorption edge of the reference spectrum. This is a **recommended**  
 295     parameter for use in an *XDI* file containing a reference spectrum.  
 296     – *Units*: none  
 297     – *Format*: same as `Element.symbol`

298 • **Namespace:** `Element` – **Tag:** `ref_edge`

- 299     – *Description*: The measured edge of the reference spectrum. This is a **recommended**  
 300     parameter for use in an *XDI* file containing a reference spectrum.  
 301     – *Units*: none  
 302     – *Format*: same as `Element.edge`

## 303     2.12 Defined items in the Column namespace

304     Items in the Column namespace describe single columns of the data table. The first column **must**  
 305     be the energy.

306     All tags in the `Column` namespace **must** be integers.

307 • **Namespace:** `Column` – **Tag:** `1`

- 308     – *Description*: A description of the abscissa array for the measured data. This is **recom-**  
 309     **mended** for use in an *XDI* file.  
 310     – *Units*: eV (**recommended**), keV, pixel, angle in degrees, angle in radians, steps  
 311     – *Format*: word + units

312 • **Namespace:** `Column` – **Tag:** `N`

- 
- 313    - *Description*: A description of the Nth column (where  $\boxed{N}$  is an integer) of the measured  
 314    data. This is **recommended** for use in an *XDI* file.  
 315    - *Units*: as needed  
 316    - *Format*: word (+ units)

317    The following labels are defined for common array types.  $\boxed{\text{Column.N}}$  items **must** use these labels  
 318    when appropriate. The array label line at the beginning of the data section of the *XDI* file also **must**  
 319    use these labels when those columns are present.

Column label	Meaning	choice of units (if required)
<code>energy</code>	mono energy	eV / keV / pixel
<code>angle</code>	mono angle	degrees / radians / steps
<code>i0</code>	monitor intensity	
<code>itrans</code>	transmission intensity	
<code>ifluor</code>	fluorescence intensity	
<code>irefer</code>	reference intensity	
<code>mutrans</code>	mu transmission	
<code>mufluor</code>	mu fluorescence	
<code>murefer</code>	mu reference	
<code>normtrans</code>	normalized mu transmission	
<code>normfluor</code>	normalized mu fluorescence	
<code>normrefer</code>	normalized mu reference	
<code>k</code>	wavenumber	
<code>chi</code>	EXAFS	
<code>chi_mag</code>	magnitude of Filtered $\chi(k)$	
<code>chi_ph</code>	phase of Filtered $\chi(k)$	
<code>chi_re</code>	real part of Filtered $\chi(k)$	
<code>chi_im</code>	imaginary part of Filtered $\chi(k)$	
<code>r</code>	radial distance	
<code>chir_mag</code>	magnitude of FT[ $\chi(k)$ ]	
<code>chir_ph</code>	phase of FT[ $\chi(k)$ ]	
<code>chir_re</code>	real part of FT[ $\chi(k)$ ]	
<code>chir_im</code>	imaginary part of FT[ $\chi(k)$ ]	

---

320    A column containing some other measurement **must** be identified with units when appropriate. For  
 321    example, a column counting time since the `Scan.start_time` timestamp might be labeled as

---

```

322      # Column.N: elapsed_time seconds

323      while a column containing an ongoing measure of temperature as a voltage on a thermocouple might
324      be labeled as

325      # Column.N: thermocouple millivolts

```

## 326    2.13 Extension fields

327    Metadata tags carry syntax and may carry semantics. That is, it is possible to have syntactically  
 328    correct tags that have no definition. Such tags could carry information considered useful by the  
 329    user or the author of software that, at some point, touches the data.

330    Such a tag could be an extension within an existing namespace. This has already been discussed in  
 331    the context of the `Sample` and `Scan` namespaces.

332    Such a tag could also be part of a new namespace. One application of a new namespace would be to  
 333    tie a group of metadata tags to a particular application. For example, the data processing program  
 334    Athena might attach tags associated with the parameters for normalizing the data. That might look  
 335    something like this:

```

336      # Athena.pre1: -150
337      # Athena.pre2: -30
338      # Athena.nor1: 150
339      # Athena.nor2: 800

```

340    These define the boundaries of the pre- and post-edge lines used to determine the edge step of the  
 341     $\mu(E)$  spectrum.

342    The use of such extension tags is encouraged for authors of controls, data acquisition, data analysis,  
 343    and data archiving software.

344    If an extension tag is not understood due its lack of defined semantics, the **recommended** behavior  
 345    for software touching the data is to silently preserve the metadata.