

XSTAR

XML System for Textual and Archaeological Research

The following XML markup language defines a database structure (schema) based on the “Extensible Markup Language (XML) 1.0” specification of the World Wide Web Consortium (www.w3.org/xml). This XML database structure and related Java software constitute the “XML System for Textual and Archaeological Research” (XSTAR). The goal of the XSTAR project is to create a sophisticated Internet-based research environment for specialists in textual and archaeological studies. In particular, XSTAR is intended for archaeologists, philologists, historians, and historical geographers who work with ancient texts, artifacts, and geographical or environmental data. The XSTAR database structure and related software are purposely generic; they are not restricted to a particular region or period.

XSTAR employs a hierarchical data model which is completely generalizable and extensible. This data model is more appropriate for spatially organized and highly variable archaeological and geographical data than are the tabular data models commonly used in archaeological research. Moreover, XSTAR makes very effective use of the hierarchical structure of XML itself to represent more directly and efficiently than relational tables could both the spatial hierarchy of archaeological units of observation (represented as **unit** elements) and logical hierarchies of archaeological variables and values (represented as **var** and **val** elements). The resulting XML tagging scheme or “markup language” has been dubbed “ArchaeoML” (Archaeological Markup Language). This approach is explained in “Archaeological Data Models and Web Publication Using XML” by David Schloen in *Computers and the Humanities*, vol. 35, May 2001.

In addition to archaeological and geographical descriptions, XSTAR describes the epigraphic and linguistic features of ancient texts (represented as **text** elements), whose physical contexts and description as artifacts, including photographs and drawings, are treated as archaeological data. The link between archaeological data and textual data is accomplished by linking the relevant archaeological **unit** element to the appropriate **text** element. Recursive tree structures (in which elements are nested within the same type of element) are used not just for the spatial hierarchy of archaeological units but also for text descriptions because: (1) tree structures are appropriate for both epigraphic and linguistic analyses of larger text components into subcomponents; (2) the number of different element types is kept to a minimum and recursive programming techniques can be used, simplifying software development; and (3) tree structures are easily implemented in XML.

Each **text** element contains two subtrees: one represents the hierarchy of physical parts that make up the text (**part** elements) and the graphic signs within those parts (**sign** elements); the other is a hierarchy of discourse units (represented as **dis** elements), however these be defined. A discourse unit can contain an alphabetic transcription and/or a modern-language translation at the appropriate hierarchical level, such that a complete transcription or translation of the text can be recursively concatenated by traversing the component discourse units. Beyond the individual text, each text in turn belongs to a corpus of texts (the **texts** element), and all texts ultimately belong to a single **texts** element which is the root of a tree structure that contains all **text** elements.

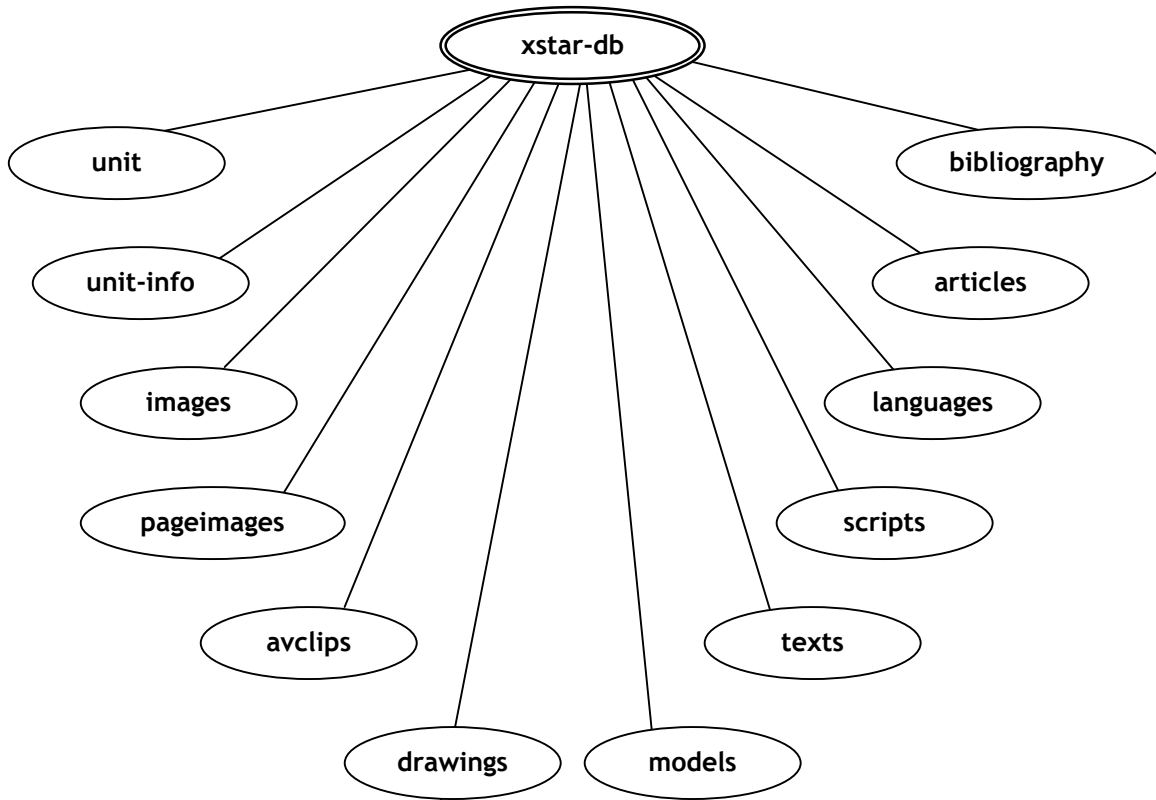
The grammatical and lexical categorization of discourse units is represented by links between the relevant **dis** element and various elements that represent a language description; i.e., the phonology, morphology, syntax, and lexicon of each ancient language in the database. For each word in a language's lexicon (represented by the **lemma** element), there is an explanatory discussion, a dictionary article whose presentation characteristics are determined by an XSL stylesheet. The dictionary article contains links to ancient texts in which the word occurs (represented as **text** elements), and to more general secondary articles or relevant bibliographic references contained in **article** and **biblio** elements (whose presentation characteristics are specified in XSL stylesheets). These articles and bibliographic references in turn can have links to primary archaeological and textual data and language descriptions.

Secondary articles and reports are organized in named sets or hierarchies, thus organizing primary archaeological and philological data, language descriptions, secondary literature, and bibliographic references under various topical headings. These reflect modern terms and concepts, as opposed to ancient languages and linguistic categories. Any number of topical rubrics may be defined, some of which will be organized in conceptual hierarchies similar to those found in traditional printed encyclopedias. These rubrics and their hierarchies can be straightforwardly represented by XML elements and element hierarchies.

The diagram below shows the hierarchical relationships and links among XSTAR element types. After that is a chart listing the names of all the element types and the names and types of the attributes for each element type. The main design principles followed here are that the number of element types should be kept to a minimum and nesting of the same element type within itself should be used wherever possible. This allows recursive techniques to be used in software that manipulates XML-structured data, and it allows the software to be as generic as possible, making it simpler and thus easier to develop.

The XSTAR Element Hierarchy

Children of **xstar-db** root element (top-level elements):



Document root element type



Occurs zero or one time



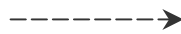
Occurs zero or more times



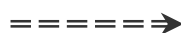
Occurs one or more times



parent-child relationship

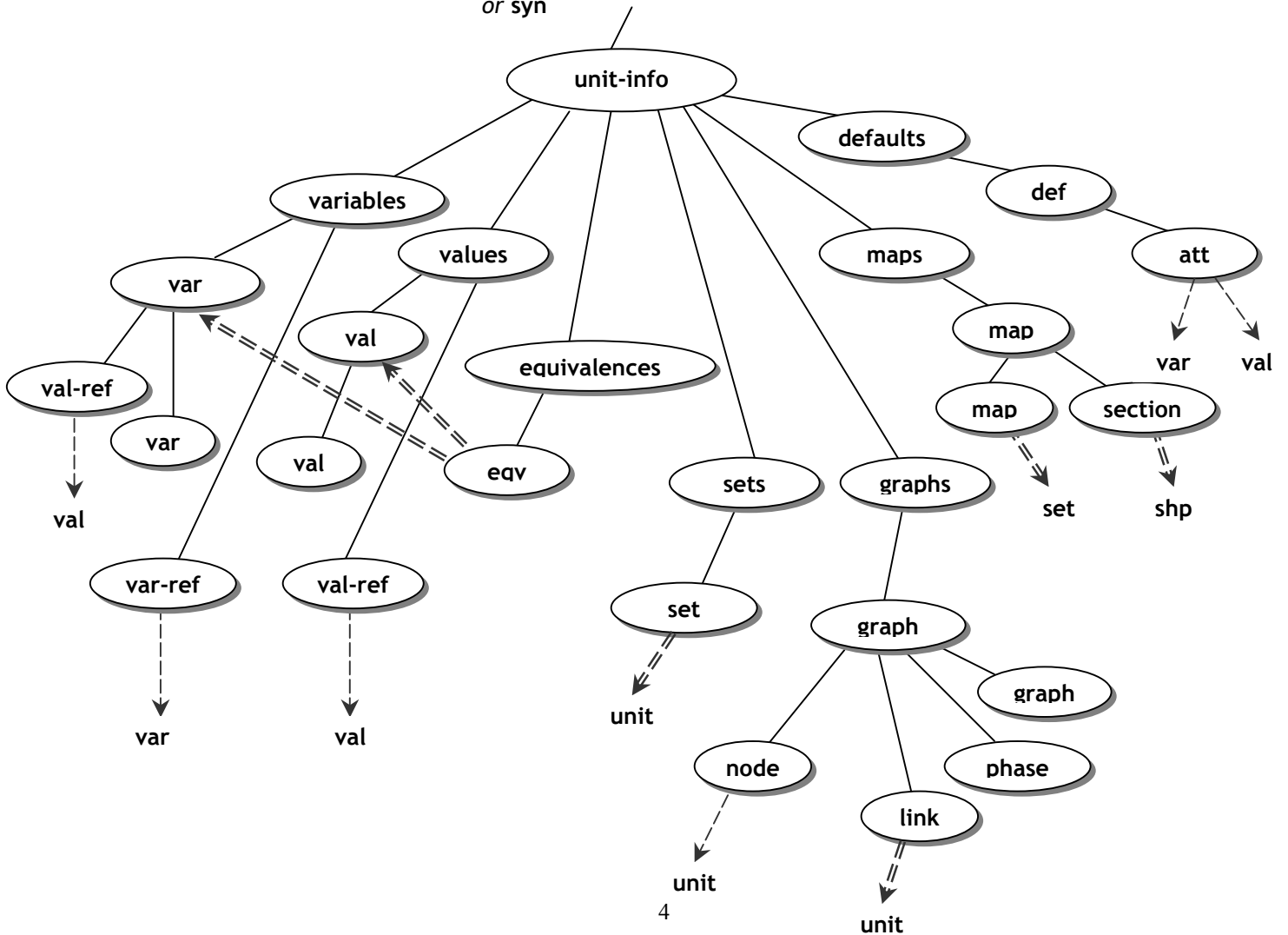
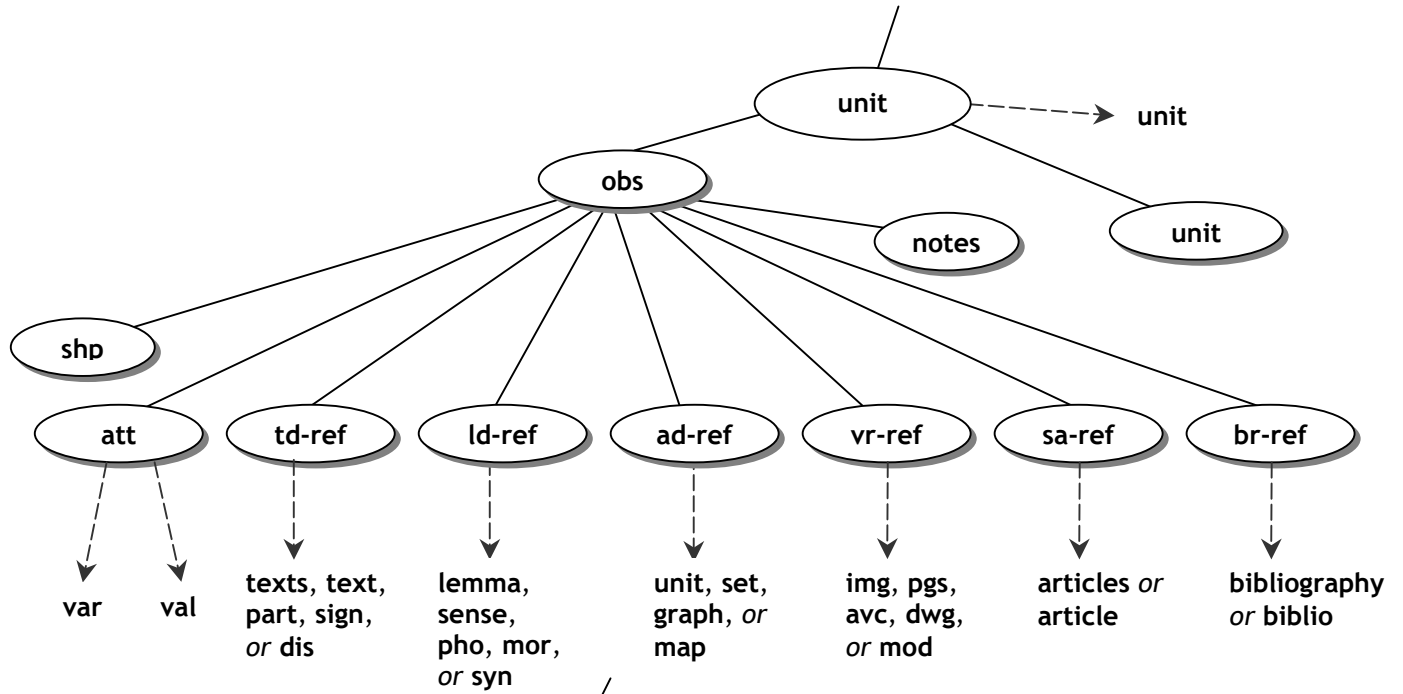


IDREF link (one element)

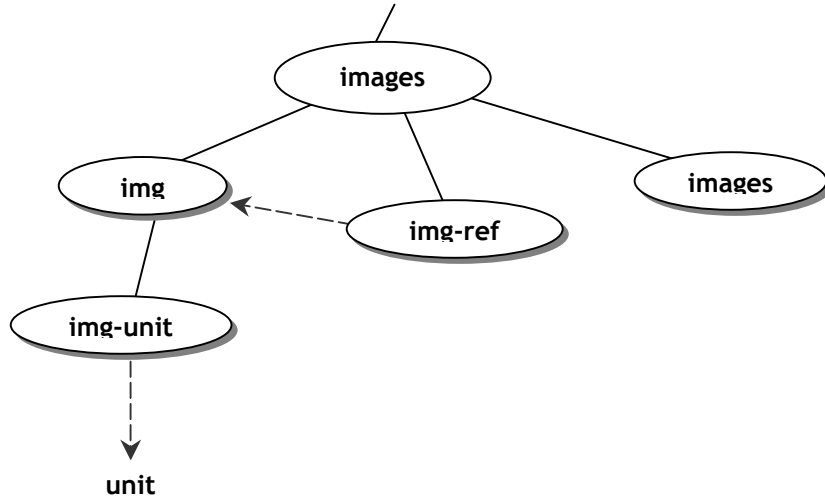


IDREFS link (more than one element)

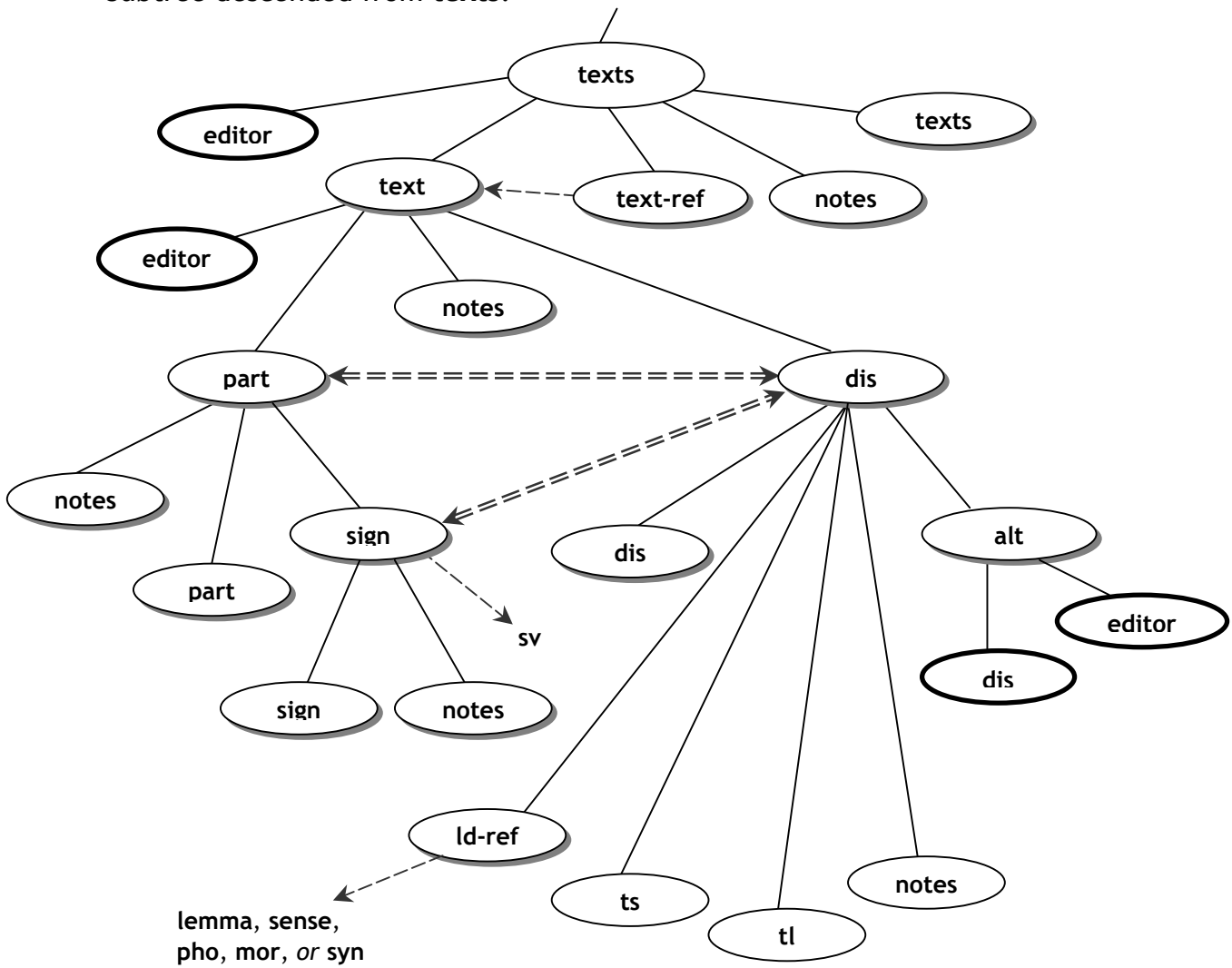
Subtrees descended from **unit** and **unit-info** elements:



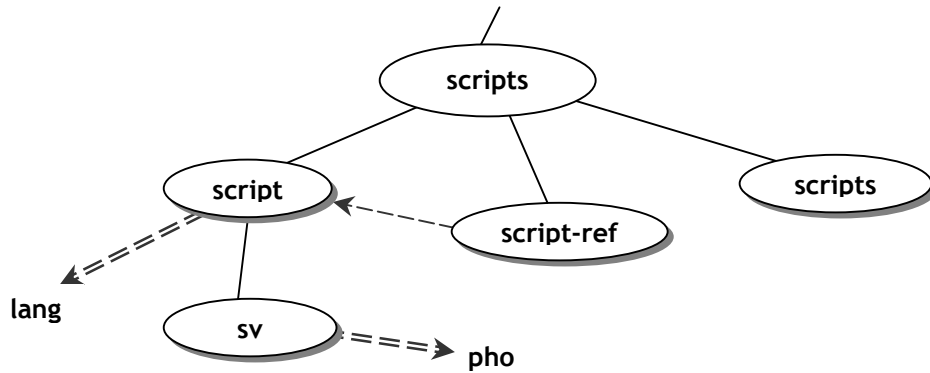
Subtree descended from **images** (same pattern for **pageimages**, **avclips**, **drawings**, and **models**):



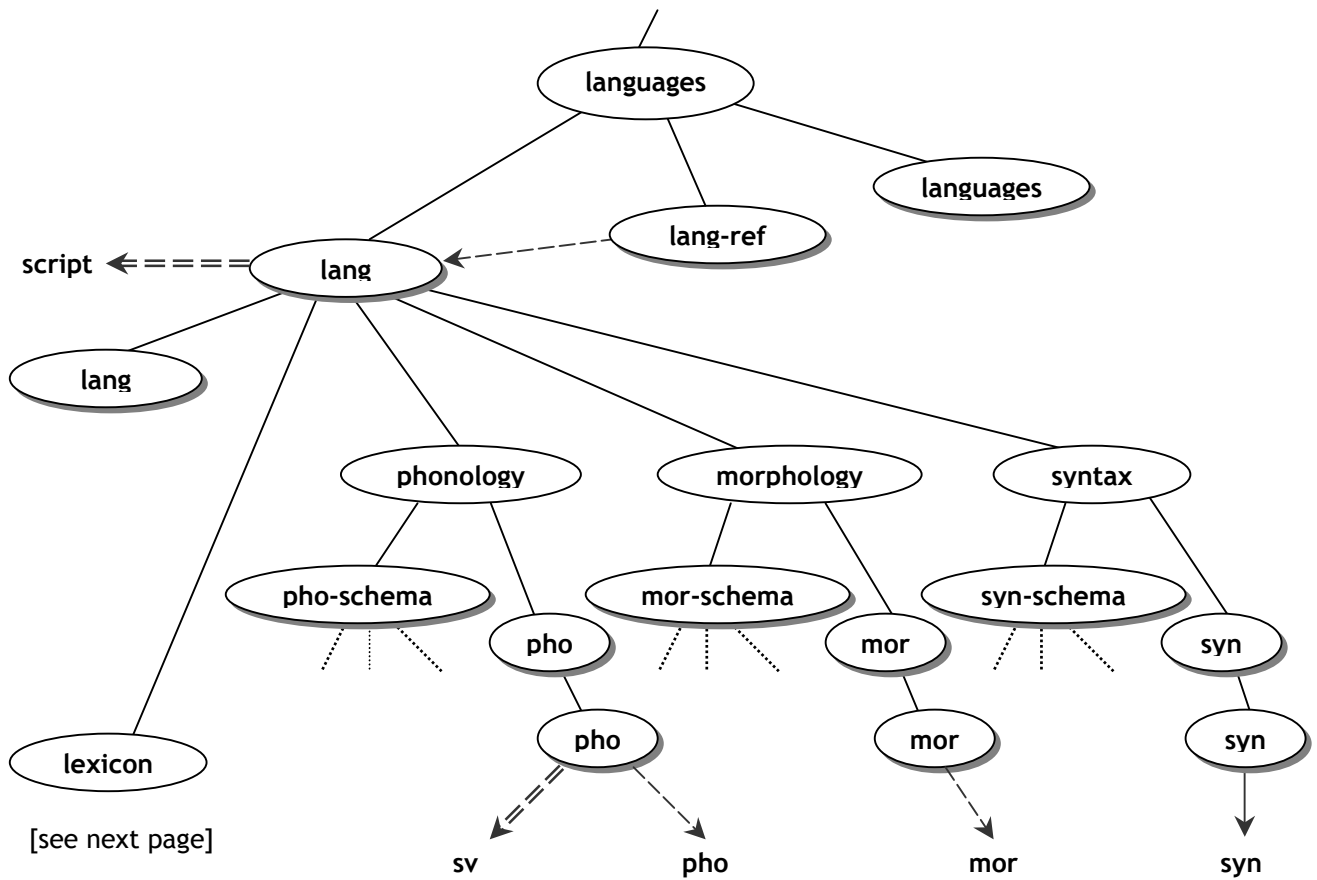
Subtree descended from **texts**:

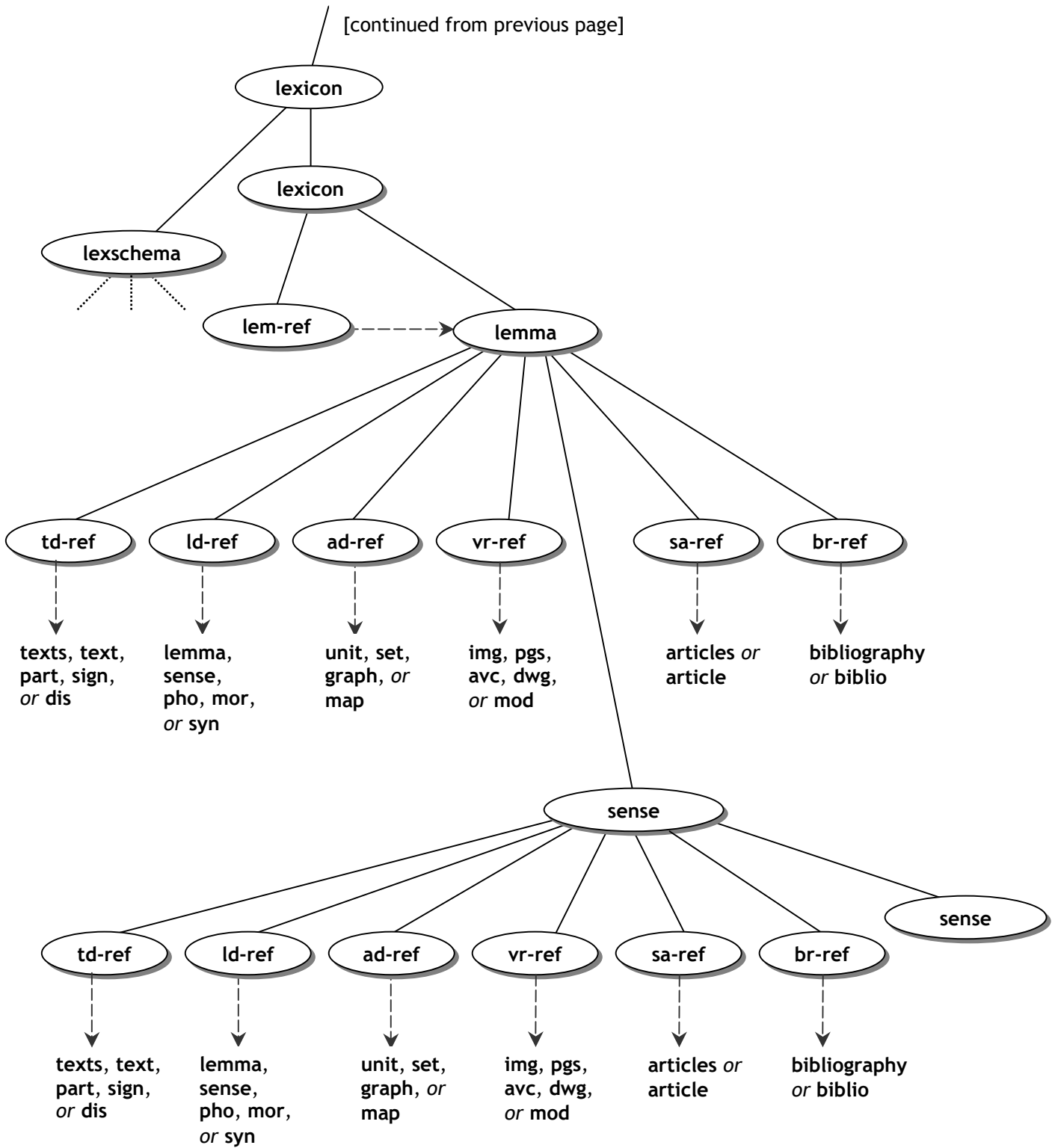


Subtree descended from **scripts**:

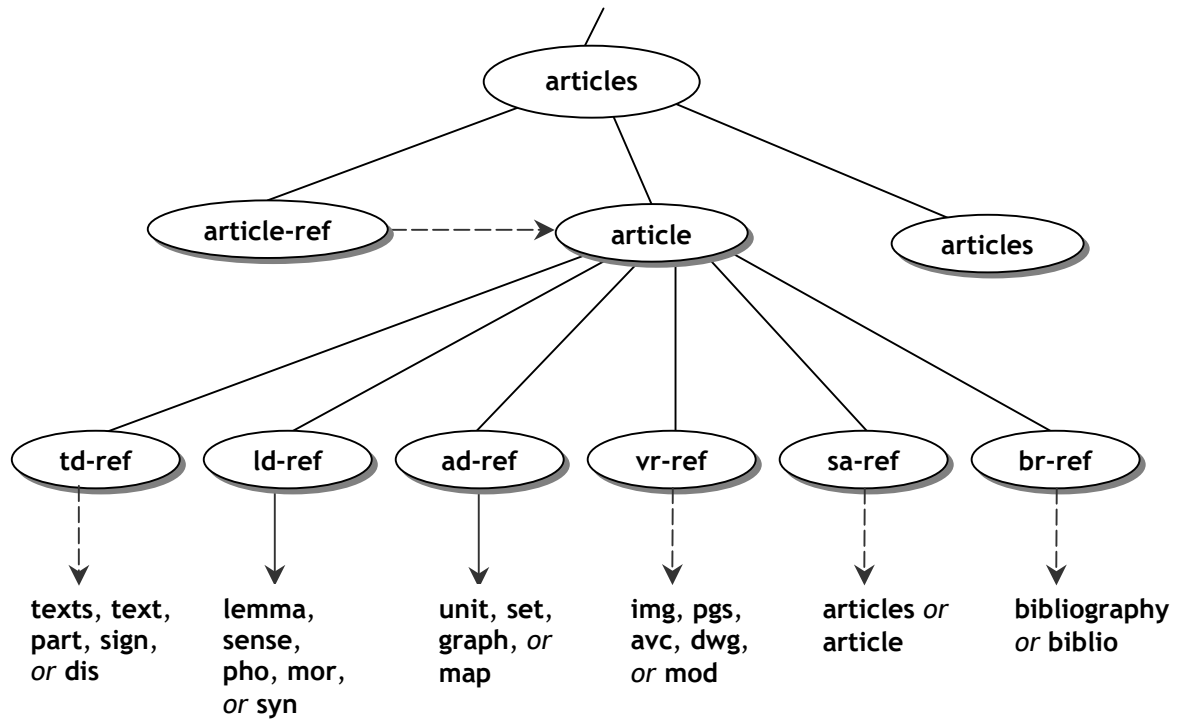


Subtree descended from **languages**:

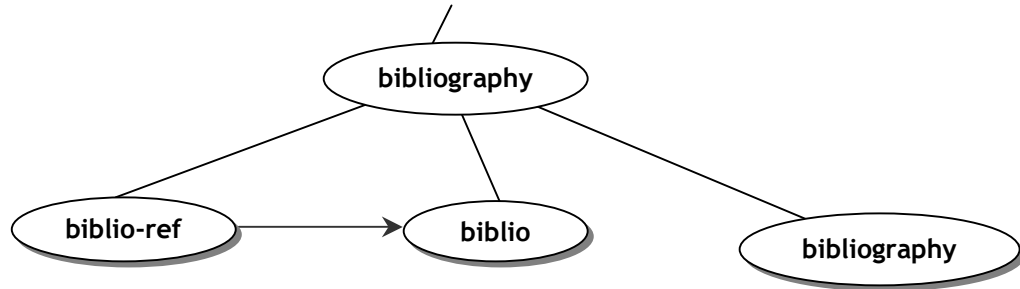




Subtree descended from **articles**:



Subtree descended from **bibliography**:



The XSTAR Markup Language

Root element type:

xstar-db

Children of xstar-db:

unit = a spatially situated unit of observation
unit-info = parent of subtrees for variables, values, equivalences, sets, graphs, and maps
images = the set of all single-page raster images, usually scanned photographs (JPEG)
pageimages = the set of all multi-page files of raster images, usually archival documents (PDF)
avclips = the set of all audio or video clips
drawings = the set of all 2D vector drawings, including animated drawings (SVG)
models = the set of all 3D vector models, including “virtual reality” models (VRML)
texts = the set of all ancient texts
scripts = the set of all ancient scripts
languages = the set of all ancient languages
articles = the set of all secondary articles and prose reports
bibliography = the set of all bibliographic references

Children of unit:

obs = an observation of the parent unit
unit = a hierarchically nested unit spatially contained within the parent unit

Children of obs:

att = an attribute of the parent observation, consisting of a variable-value pair
shp = a map shape associated with the parent observation of the grandparent unit
td-ref = a reference to an XSTAR text-description element (**texts**, **text**, **part**, **sign**, **dis**)
ld-ref = a reference to an XSTAR language-description element (**lemma**, **sense**, **pho**, **mor**, **syn**)
ad-ref = a reference to an XSTAR archaeological-description element (**unit**, **set**, **graph**, **map**)
vr-ref = a reference to an XSTAR visual-resource element (**img**, **pgs**, **avc**, **dwg**, **mod**)
sa-ref = a reference to an XSTAR secondary-articles element (**articles**, **article**)
br-ref = a reference to an XSTAR bibliographic-references element (**bibliography**, **biblio**)
notes = notes about the parent observation

Children of unit-info:

variables = a set of variables that describe units
values = a set of values of nominal and ordinal variables
defaults = a set of named sets of default attributes to be used to describe observations of units
equivalences = a set of equivalences among variables and values
sets = a set of named sets of units of observation (e.g., query result-sets)
graphs = a set of named node-link graphs indicating interrelationships among units
maps = a set of named 2D georeferenced maps of units

Children of variables:

var = a variable that describes some unit
var-ref = a reference to a **var** element (permitting overlapping sets of variables)

Children of var:

val-ref = a reference to a **val** element
var = a hierarchically nested **var** element that is logically subordinate to the parent **var**

Children of values:

- val** = a value of a nominal or ordinal variable
- val-ref** = a reference to a **val** element (permitting overlapping sets of values)

Children of val:

- val** = a hierarchically nested **val** element that is logically subordinate to the parent **val**

Children of defaults:

- def** = a named set of default attributes that can be used to describe observations of units

Children of def:

- att** = an attribute of a potential observation of a unit, consisting of a variable-value pair

Children of equivalences:

- eqv** = an equivalence between two variables or two nominal or ordinal values

Children of sets:

- set** = a set of units (e.g., a query result-set)

Children of graphs:

- graph** = a node-link graph structure indicating interrelationships among units

Children of graph:

- node** = a node of a graph
- link** = a link of a graph
- phase** = a stratigraphic phase in a "Harris Matrix" node-link graph
- graph** = a hierarchically nested graph within the parent graph

Children of maps:

- map** = a 2D georeferenced map of a set of units

Children of map:

- section** = a section drawing of units corresponding to a section line drawn on the parent map
- map** = a hierarchically nested map within the parent map

Children of images:

- img** = a raster image
- img-ref** = a reference to an **img** element (permitting overlapping sets of images)
- images** = a hierarchically nested subset of the parent set of images

Children of img:

- img-unit** = a cross-reference from an image to a unit

Children of pageimages:

- pgs** = a multi-page file of images (e.g., a scanned archival document in PDF format)
- pgs-ref** = a reference to a **pgs** element (permitting overlapping sets of multi-page files)
- pageimages** = a hierarchically nested subset of the parent file of page-images

Children of pgs:

- pgs-unit** = a cross-reference from a file of page-images to a unit

Children of avclips:

- avc** = an audio or video clip
- avc-ref** = a reference to an **avc** element (permitting overlapping sets of avclips)
- avclips** = a hierarchically nested subset of the parent set of avclips

Children of avc:

- avc-unit** = a cross-reference from an avclip to a unit

Children of drawings:

- dwg** = a 2D vector drawing
- dwg-ref** = a reference to a **dwg** element (permitting overlapping sets of drawings)
- drawings** = a hierarchically nested subset of the parent set of 2D vector drawings

Children of dwg:

- dwg-unit** = a cross-reference from a 2D vector drawing to a unit

Children of models:

- mod** = a 3D vector model
- mod-ref** = a reference to a **mod** element (permitting overlapping sets of models)
- models** = a hierarchically nested subset of the parent set of 3D vector models

Children of mod:

- mod-unit** = a cross-reference from a 3D vector model to a unit

Children of texts:

- text** = an ancient text
- text-ref** = a reference to a **text** element (permitting overlapping sets of texts)
- editor** = the editor(s) of the parent set of texts
- notes** = notes about the parent set of texts
- texts** = a hierarchically nested subset of the parent set of texts

Children of text:

- part** = a physical part of the parent text
- dis** = a discourse unit within the parent text
- editor** = the editor(s) of the parent text
- notes** = notes about the parent text

Children of part:

- sign** = a graphic sign contained within the parent part
- notes** = notes about the parent part
- part** = a hierarchically nested subpart contained within the parent part

Children of sign:

- notes** = notes about the parent sign
- sign** = a hierarchically nested sign contained within the parent sign

Children of dis:

- ld-ref** = a reference to an XSTAR language-description element (**lemma, sense, pho, mor, syn**)
- ts** = a normalized Latin-alphabet transcription of the parent discourse unit
- tl** = a modern-language translation of the parent discourse unit
- notes** = notes about the parent discourse unit
- alt** = an alternative interpretation of the parent discourse unit
- dis** = a hierarchically nested discourse unit contained within the parent discourse unit

Children of alt:

- dis** = a discourse unit specified in the parent interpretation
- editor** = the editor(s) of the parent interpretation

Children of scripts:

- script** = an ancient script
- script-ref** = a reference to a **script** element (permitting overlapping sets of scripts)
- scripts** = a hierarchically nested subset of the parent set of scripts

Children of script:

- sv** = a sign-value pair belonging to the parent script

Children of languages:

- lang** = an ancient language
- lang-ref** = a reference to a **lang** element (permitting overlapping sets of languages)
- languages** = a hierarchically nested subset of the parent set of languages

Children of lang:

- phonology** = the phonology of the parent language
- morphology** = the morphology of the parent language
- syntax** = the syntax of the parent language
- lexicon** = the lexicon of the parent language
- lang** = a hierarchically nested sublanguage within the parent language

Children of phonology:

- pho** = a phonological element
- pho-schema** = a phonological schema

Children of pho:

- pho** = a hierarchically nested phonological element

Children of morphology:

- mor** = a morphological element
- mor-schema** = a morphological schema

Children of mor:

- mor** = a hierarchically nested morphological element

Children of syntax:

- syn** = a syntactic element
- syn-schema** = a syntactic schema

Children of syn:

syn = a hierarchically nested syntactic element

Children of lexicon:

lemma = a lexical lemma

lem-ref = a reference to a **lemma** element (permitting overlapping sets of lemmas)

lex-schema = contains XSL specifying the format for presenting dictionaries of a given type

lexicon = a hierarchically nested sublexicon of the parent lexicon

Children of lemma:

sense = a dictionary sense of the parent lemma

td-ref = a reference to an XSTAR text-description element (**texts, text, part, sign, dis**)

ld-ref = a reference to an XSTAR language-description element (**lemma, sense, pho, mor, syn**)

ad-ref = a reference to an XSTAR archaeological-description element (**unit, set, graph, map**)

vr-ref = a reference to an XSTAR visual-resource element (**img, pgs, avc, dwg, mod**)

sa-ref = a reference to an XSTAR secondary-articles element (**articles, article**)

br-ref = a reference to an XSTAR bibliographic-references element (**bibliography, biblio**)

Children of sense:

td-ref = a reference to an XSTAR text-description element (**texts, text, part, sign, dis**)

ld-ref = a reference to an XSTAR language-description element (**lemma, sense, pho, mor, syn**)

ad-ref = a reference to an XSTAR archaeological-description element (**unit, set, graph, map**)

vr-ref = a reference to an XSTAR visual-resource element (**img, pgs, avc, dwg, mod**)

sa-ref = a reference to an XSTAR secondary-articles element (**articles, article**)

br-ref = a reference to an XSTAR bibliographic-references element (**bibliography, biblio**)

sense = a hierarchically nested subsense of the parent sense

Children of articles:

article = a secondary article or prose report

article-ref = a reference to an **article** element (permitting overlapping sets of articles)

articles = a hierarchically nested subset of the parent set of articles

Children of article:

td-ref = a reference to an XSTAR text-description element (**texts, text, part, sign, dis**)

ld-ref = a reference to an XSTAR language-description element (**lemma, sense, pho, mor, syn**)

ad-ref = a reference to an XSTAR archaeological-description element (**unit, set, graph, map**)

vr-ref = a reference to an XSTAR visual-resource element (**img, pgs, avc, dwg, mod**)

sa-ref = a reference to an XSTAR secondary-articles element (**articles, article**)

br-ref = a reference to an XSTAR bibliographic-references element (**bibliography, biblio**)

Children of bibliography:

biblio = a bibliographic reference

biblio-ref = a reference to a **biblio** element (permitting overlapping sets of bibliographic references)

bibliography = a hierarchically nested subset of the parent bibliography

XSTAR Element Types and Attributes

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
xstar-db	name	CDATA	yes	database name
	author	CDATA	no	database author(s)
	date	CDATA	no	creation date
	revisions	CDATA	no	revision history
<i>(N.B.: The database description, if any, is stored as the character data content of the xstar-db element. Note that the use of the term “database” is not meant to imply that a single XSTAR document, as a logical structure, is contained in one physical file or one Internet location.)</i>				
unit	id	ID	yes	unique identifier (“u”+integer)
	name	CDATA	yes	unit name
	altname	CDATA	no	alternate name
	ref	IDREF	no	points to another unit element
<i>(N.B.: If there is a ref to another unit element, the current unit element acts only as a placeholder duplicating that item; it cannot have child elements. No character data content is allowed in a unit element, i.e., no mixed content.)</i>				
obs	name	CDATA	yes	observation name
	def	enumerated	yes	default observation? (yes/no)
	date	CDATA	no	date of the observation
	author	CDATA	no	author(s)/recorder(s) of the observation
	defimg	IDREF	no	points to the default img element
att	varref	IDREF	yes	points to a var element
	valref	IDREF	no	points to a val element
<i>(N.B.: If there is no valref attribute, the character data content of the att element is the variable’s value; this will be the case if the referenced var element is of type “int,” “dec,” “cal,” or “anum.”)</i>				
notes	date	CDATA	no	date of the notes
	lang	enumerated	yes	language of the notes
	author	CDATA	no	author(s) of the notes
<i>(N.B.: The notes are stored as the character data content of the notes element.)</i>				
shp	id	ID	yes	unique identifier (“s”+integer)
	type	enumerated	yes	shape type (point/polyline/polygon)
	vert	NMTOKENS	yes	shape vertices (usu. UTM coordinates)
	zone	NMTOKEN	yes	UTM zone
	sect	enumerated	yes	section shape? (yes/no; defaults to no)
	elev	NMTOKEN	no	shape elevation in meters above MSL
	style	NMTOKENS	no	overrides class style
	ord	enumerated	no	draw order (front/back/any; defaults to any)
	spot	enumerated	no	spot shape (elev/grid/label/no; defaults to no)
	label	enumerated	no	show item name on shape? (yes/no; defaults to no)
	annot	enumerated	no	show annotation? (yes/no; defaults to no)
	<i>(N.B.: If spot is not “no” then type must be “point.” If an annotation is associated with a shape, this is stored as the character data content of the shp element.)</i>			
td-ref	ref	IDREF	yes	points to a texts , text , part , sign , or dis element
	type	enumerated	yes	type of text-description element referenced
<i>(N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the td-ref element.)</i>				
ld-ref	ref	IDREF	yes	points to lemma , sense , pho , mor , syn element
	type	enumerated	yes	type of language-description element referenced
<i>(N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the ld-ref element.)</i>				

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
ad-ref	ref	IDREF	yes	<i>points to a unit, set, graph, or map element type of archaeological-desc. element ref'd (N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the ad-ref element.)</i>
	type	enumerated	yes	
vr-ref	ref	IDREF	yes	<i>points to an img, pgs, avc, dwg, or mod element type of visual-resource element referenced (N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the vr-ref element.)</i>
	type	enumerated	yes	
sa-ref	ref	IDREF	yes	<i>points to an articles or article element type of secondary-article element referenced (N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the sa-ref element.)</i>
	type	enumerated	yes	
br-ref	ref	IDREF	yes	<i>points to a bibliography or biblio element type of bibliographic-ref. element ref'd (N.B.: If an annotation is associated with the element reference, this is stored as the character data content of the br-ref element.)</i>
	type	enumerated	yes	
unit-info	name	CDATA	no	<i>name of unit-info author(s)</i>
	author	CDATA	no	
variables	name	CDATA	yes	<i>name of variable set (N.B.: The variable set description, if any, is stored as the character data content of the variables element.)</i>
var	id	ID	yes	<i>unique identifier (“var”+integer) variable name variable type (nom/ord/log/int/dec/cal/anum) abbreviated variable name (N.B.: Possible var types are: nominal (nom), ordinal (ord), logical (log), integral (int), decimal (dec), calendrical (cal), alphanumeric (anum). The order of the id’s listed in vals is significant if type is “ord.” Free-form notes pertaining to the variable, if any, are stored as the character data content of the var element.)</i>
	name	CDATA	yes	
	type	enumerated	yes	
	abbr	CDATA	no	
var-ref	ref	IDREF	yes	<i>points to a var element</i>
values	name	CDATA	yes	<i>name of value set (N.B.: The value set description, if any, is stored as the character data content of the values element.)</i>
val	id	ID	yes	<i>unique identifier (“val”+integer) actual nominal, ordinal, or logical value abbreviated value name (N.B.: Free-form notes pertaining to the value, if any, are stored as the character data content of the val element.)</i>
	name	CDATA	yes	
	abbr	CDATA	no	
val-ref	ref	IDREF	yes	<i>points to a val element</i>
defaults	name	CDATA	yes	<i>name of set of named set of defaults (N.B.: The defaults set description, if any, is stored as the character data content of the defaults element.)</i>
def	name	CDATA	yes	<i>name of a set of default attributes</i>

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
equivalences	name	CDATA	yes	<i>name of equivalence set</i>
	<i>(N.B.: The equivalence set description, if any, is stored as the character data content of the equivalences element.)</i>			
eqv	refs	IDREFS	yes	<i>points to two var elements or two val elements</i>
	prob	NMTOKEN	no	<i>probability of equivalence</i>
	deg	NMTOKEN	no	<i>degree of equivalence</i>
sets	name	CDATA	yes	<i>name of set of unit-sets</i>
	<i>(N.B.: The description of the set of unit-sets, if any, is stored as the character data content of the sets element.)</i>			
set	id	ID	yes	<i>unique identifier (“set”+integer)</i>
	name	CDATA	yes	<i>name of unit-set</i>
	criteria	CDATA	yes	<i>query criteria for selecting units in the set</i>
	style	NMTOKENS	yes	<i>line style, fill pattern, color, etc.</i>
	rundatetime	NMTOKENS	yes	<i>last date and time set was refreshed</i>
	numberofunits	NMTOKEN	yes	<i>number of units in set</i>
	units	IDREFS	yes	<i>points to one or more unit elements</i>
	<i>(N.B.: The unit-set description, if any, is stored as the character data content of the set element. **Note that the format for specifying set criteria, making use of unit descriptions plus tree, graph, and map relationships, is not given here.**)</i>			
graphs	name	CDATA	yes	<i>name of graph set</i>
	<i>(N.B.: The graph set description, if any, is stored as the character data content of the graphs element.)</i>			
graph	id	ID	yes	<i>unique identifier (“graph”+integer)</i>
	name	CDATA	yes	<i>name of node-link graph</i>
	type	<i>enumerated</i>	yes	<i>graph type (sequence/adjacency)</i>
	pos	NMTOKENS	no	<i>nested position (node pos.) within a parent graph</i>
	<i>(N.B.: The graph description, if any, is stored as the character data content of the graph element.)</i>			
node	unit	IDREF	yes	<i>points to a unit element</i>
	pos	NMTOKENS	yes	<i>position of unit in parent graph (col, row)</i>
<i>(N.B.: Free-form notes pertaining to the node, if any, are stored as the character data content of the node element.)</i>				
link	units	IDREFS	yes	<i>points to two linked unit elements</i>
	type	<i>enumerated</i>	yes	<i>link type (seq/corr/adj)</i>
	pos	NMTOKENS	yes	<i>start and end positions in parent graph</i>
<i>(N.B.: Free-form notes pertaining to the link, if any, are stored as the character data content of the link element.)</i>				
phase	name	CDATA	yes	<i>phase name</i>
	pos	NMTOKENS	yes	<i>start and end rows in parent graph</i>
<i>(N.B.: The phase description, if any, is stored as the character data content of the phase element.)</i>				
maps	name	CDATA	yes	<i>name of map set</i>
	<i>(N.B.: The map set description, if any, is stored as the character data content of the maps element.)</i>			

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
map	id	ID	yes	unique identifier (“map”+integer)
	name	CDATA	yes	map name
	sets	IDREFS	yes	points to one or more set elements
	trans	NMTOKENS	no	transformation params. to local coords.
	key	CDATA	no	map key information
	pos	NMTOKENS	no	nested position within a parent map
<i>(N.B.: The map description, if any, is stored as the character data content of the map element.)</i>				
section	name	CDATA	yes	section name
	shps	IDREFS	yes	points to one or more shp elements
	pos	NMTOKENS	yes	start and end coords. within parent map
	vrange	NMTOKENS	yes	top and bottom elevations of section
<i>(N.B.: The section description, if any, is stored as the character data content of the section element.)</i>				
images	name	CDATA	yes	name of raster image set
	<i>(N.B.: The image set description, if any, is stored as the character data content of the images element.)</i>			
img	id	ID	yes	unique identifier (“img”+integer)
	name	CDATA	yes	raster image name
	type	enumerated	yes	image format (e.g., jpeg/tiff/gif)
	src	CDATA	yes	URI to binary image data
<i>(N.B.: The image description, if any, is stored as the character data content of the img element.)</i>				
img-ref	ref	IDREF	yes	points to an img element
img-unit	unit	IDREF	yes	points to a unit element
	pos	NMTOKENS	yes	position (coordinates) of unit in image
	label	enumerated	yes	show unit name in image? (yes/no; def. no)
	annot	enumerated	no	show annotation? (yes/no; defaults to no)
<i>(N.B.: If an annotation is associated with a unit’s hotspot on an image, this is stored as the character data content of the img-unit element.)</i>				
pageimages	name	CDATA	yes	name of set of page-images
	<i>(N.B.: The page-image set description, if any, is stored as the character data content of the pageimages element.)</i>			
pgs	id	ID	yes	unique identifier (“pgs”+integer)
	name	CDATA	yes	page-images name
	type	enumerated	yes	image format (e.g., pdf)
	src	CDATA	yes	URI to binary page-images data
<i>(N.B.: The page-images description, if any, is stored as the character data content of the pgs element.)</i>				
pgs-ref	ref	IDREF	yes	points to an pgs element
pgs-unit	unit	IDREF	yes	points to a unit element
	pos	NMTOKENS	yes	position (coordinates) of unit in page-images
	label	enumerated	yes	show unit name in page-images? (yes/no; def. no)
	annot	enumerated	no	show annotation? (yes/no; defaults to no)
<i>(N.B.: If an annotation is associated with a unit’s hotspot on a page-images display, this is stored as the character data content of the pgs-unit element.)</i>				

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
avclips	name	CDATA	yes	<i>name of audio/video clip set</i> (<i>N.B.: The audio/video clip set description, if any, is stored as the character data content of the avclips element.</i>)
avc	id	ID	yes	<i>unique identifier (“avc”+integer)</i>
	name	CDATA	yes	<i>audio/video clip name</i>
	type	<i>enumerated</i>	yes	<i>audio/video format (e.g., mpeg/avi)</i>
	src	CDATA	yes	<i>URI to binary audio/video data</i> (<i>N.B.: The audio/video clip description, if any, is stored as the character data content of the avc element.</i>)
avc-ref	ref	IDREF	yes	<i>points to an avc element</i>
avc-unit	unit	IDREF	yes	<i>points to a unit element</i>
	pos	NMTOKENS	yes	<i>position of unit in audio/video clip</i>
	label	<i>enumerated</i>	yes	<i>show unit name in clip? (yes/no; def. no)</i>
	annot	<i>enumerated</i>	no	<i>show annotation? (yes/no; defaults to no)</i> (<i>N.B.: If an annotation is associated with a unit’s hotspot in an audio/video clip, this is stored as the character data content of the avc-unit element.</i>)
drawings	name	CDATA	yes	<i>name of 2D vector drawing set</i> (<i>N.B.: The drawing set description, if any, is stored as the character data content of the drawings element.</i>)
dwg	id	ID	yes	<i>unique identifier (“dwg”+integer)</i>
	name	CDATA	yes	<i>vector drawing name</i>
	type	<i>enumerated</i>	yes	<i>drawing format (e.g., dxf/dwg)</i>
	src	CDATA	yes	<i>URI to drawing data</i> (<i>N.B.: The drawing description, if any, is stored as the character data content of the dwg element.</i>)
dwg-ref	ref	IDREF	yes	<i>points to a dwg element</i>
dwg-unit	ref	IDREF	yes	<i>points to a unit element</i>
	pos	NMTOKENS	yes	<i>position (coordinates) of unit in drawing</i>
	label	<i>enumerated</i>	yes	<i>show unit name in drawing? (yes/no; def. no)</i>
	annot	<i>enumerated</i>	no	<i>show annotation? (yes/no; defaults to no)</i> (<i>N.B.: If an annotation is associated with a unit’s hotspot on a drawing, this is stored as the character data content of the dwg-unit element.</i>)
models	name	CDATA	yes	<i>name of 3D model set</i> (<i>N.B.: The 3D model set description, if any, is stored as the character data content of the models element.</i>)
mod	id	ID	yes	<i>unique identifier (“mod”+integer)</i>
	name	CDATA	yes	<i>3D model name</i>
	type	<i>enumerated</i>	yes	<i>3D model format</i>
	src	CDATA	yes	<i>URI to binary model data</i> (<i>N.B.: The 3D model description, if any, is stored as the character data content of the mod element.</i>)
mod-ref	ref	IDREF	yes	<i>points to a mod element</i>
mod-unit	unit	IDREF	yes	<i>points to a unit element</i>
	pos	NMTOKENS	yes	<i>position of unit in 3D model</i>
	label	<i>enumerated</i>	yes	<i>show unit name in model? (yes/no; def. no)</i>
	annot	<i>enumerated</i>	no	<i>show annotation? (yes/no; defaults to no)</i> (<i>N.B.: If an annotation is associated with a unit’s hotspot in a 3D model, this is stored as the character data content of the mod-unit element.</i>)

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
texts	id	ID	yes	unique identifier (“texts”+integer)
	name	CDATA	yes	name of text corpus
	type	<i>enumerated</i>	no	type of text corpus
	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements
	units	IDREFS	no	points to unit elements
text	id	ID	yes	unique identifier (“text”+integer)
	name	CDATA	yes	text name
	type	<i>enumerated</i>	no	text genre or category
	rfont	<i>enumerated</i>	no	Latin-based SVG font for transcriptions
	tfont	<i>enumerated</i>	no	Latin-based SVG font for translations
	lang	IDREF	no	points to a lang element
	script	IDREF	no	points to a script element
	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements
	units	IDREFS	no	points to unit elements
text-ref	ref	IDREF	yes	points to a text element
editor	id	ID	yes	unique identifier (“ed”+integer)
	name	CDATA	yes	name of editor(s)
	date	CDATA	no	date of this edition or interpretation
part	id	ID	yes	unique identifier (“p”+integer)
	type	<i>enumerated</i>	yes	part type (line/col/obv/rev/edge/etc.)
	name	CDATA	no	part name
	con	<i>enumerated</i>	no	physical condition (damaged/erased/etc.)
	dis	IDREFS	no	points to one or more dis elements
	script	IDREF	no	points to a script element
	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements
	units	IDREFS	no	points to unit elements
sign	id	ID	yes	unique identifier (“s”+integer)
	sv	IDREF	yes	points to an sv element
	con	<i>enumerated</i>	no	physical condition (damaged/erased/etc.)
	err	<i>enumerated</i>	no	scribal error (ditto/omit)
	dis	IDREFS	no	points to one or more dis elements
	script	IDREF	no	points to a script element
	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements
	units	IDREFS	no	points to unit elements
dis	id	ID	yes	unique identifier (“d”+integer)
	signs	IDREFS	no	points to one or more sign elements
	parts	IDREFS	no	points to one or more part elements
	lang	IDREF	no	points to a lang element
	lemma	IDREF	no	points to a lemma element
	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements
alt	articles	IDREFS	no	points to article elements
	biblios	IDREFS	no	points to biblio elements

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
ts	type	enumerated	yes	type of transcription
	font	enumerated	no	Latin-based font for this transcription (N.B.: The transcription is stored as the character data content of the ts element, normally using a custom font delivered in XML-SVG format.)
tl	lang	enumerated	yes	language of this translation
	type	enumerated	yes	type of translation
	font	enumerated	no	Latin-based SVG font for this translation (N.B.: The translation is stored as the character data content of the tl element, possibly using a custom font delivered in XML-SVG format.)
scripts	name	CDATA	no	name of script set
script	name	CDATA	yes	name of script
	type	enumerated	yes	type of script
	langs	IDREFS	yes	points to lang elements (langs. script used for)
	sfont	enumerated	yes	SVG sign font for displaying this script (one font character per graphic sign)
	vfont	enumerated	yes	font for transliteration of sign values
script-ref	ref	IDREF	yes	points to a script element
sv	id	ID	yes	unique identifier ("sv"+integer)
	code	CDATA	yes	standard sign number of graphic sign
	val	CDATA	yes	transliteration of one value of this sign
	phos	IDREFS	no	points to one or more pho elements (N.B.: A given sign may have more than one possible value. Each such value is represented by a different sv element storing a different val attribute for the same code .)
languages	name	CDATA	no	name of language set
lang	name	enumerated	yes	language or dialect name
	scripts	IDREFS	yes	points to script elements
lang-ref	ref	IDREF	yes	points to a lang element
phonology	editor	CDATA	yes	editor(s) of phonology information
pho-schema	author	CDATA	yes	author(s) of phonology schema
	date	CDATA	yes	date of schema (N.B.: A free-form description of the phonology schema is stored as the character data content of the pho-schema element.)
pho	id	ID	yes	unique identifier ("pho"+integer)
	cat	CDATA	yes	phonological category (e.g., "phoneme")
	val	CDATA	yes	specific value within phonological category
	svs	IDREFS	no	points to one or more sv elements
	ref	IDREF	no	cross-reference to another pho element (N.B.: If there is a ref attribute, the cat and val attributes are ignored because the pho element is then treated as a placeholder for the cross-referenced pho element.)
morphology	editor	CDATA	yes	editor(s) of morphology information
mor-schema	author	CDATA	yes	author(s) of morphology schema
	date	CDATA	yes	date of schema (N.B.: A free-form description of the morphology schema is stored as the character data content of the mor-schema element.)

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
mor	id	ID	yes	unique identifier (“mor”+integer)
	cat	CDATA	yes	morphological category
	val	CDATA	yes	specific value within morphological category
	ref	IDREF	no	cross-reference to another mor element
<i>(N.B.: If there is a ref attribute, the cat and val attributes are ignored because the mor element is then treated as a placeholder for the cross-referenced mor element.)</i>				
syntax	editor	CDATA	no	editor(s) of syntax information
syn-schema	author	CDATA	yes	author(s) of syntax schema
	date	CDATA	yes	date of schema
<i>(N.B.: A free-form description of the syntax schema is stored as the character data content of the syn-schema element.)</i>				
syn	id	ID	yes	unique identifier (“syn”+integer)
	cat	CDATA	yes	syntactic category
	val	CDATA	yes	specific value within syntactic category
	ref	IDREF	no	cross-reference to another syn element
<i>(N.B.: If there is a ref attribute, the cat and val attributes are ignored because the syn element is then treated as a placeholder for the cross-referenced syn element.)</i>				
lexicon	editor	CDATA	yes	editor(s) of lexicon or sublexicon
lex-schema	author	CDATA	yes	author(s) of dictionary schema
	date	CDATA	yes	date of schema
<i>(N.B.: A free-form description of the dictionary schema is stored as the character data content of the lex-schema element.)</i>				
lemma	id	ID	yes	unique identifier (“lemma”+integer)
	name	CDATA	yes	citation form of the word defined in the entry
	base	CDATA	no	base form of the word defined in the entry
	mor	IDREF	no	points to a mor element
<i>(N.B.: A dictionary article is stored as the character data content of the lemma element, and the various child element types can be interspersed within it. The optional base attribute is the form of the word that reflects the linguistic organizing principle of the dictionary, if this form is not identical to the citation form. The mor attribute would normally be used to indicate the basic part of speech ascribed to the word defined in the article.)</i>				
lem-ref	ref	IDREF	yes	cross-reference to a lemma element
	type	<i>enumerated</i>	yes	etymon/equivalent/synonym/antonym/other
	certain	<i>enumerated</i>	no	is the type of reference certain?(yes/no)
sense	id	ID	yes	unique identifier (“sense”+integer)
	type	<i>enumerated</i>	no	type of definition
	lang	<i>enumerated</i>	no	modern language in which definition is written
<i>(N.B.: A dictionary definition is stored as the character data content of the sense element.)</i>				
articles	name	CDATA	yes	name of article set
article	id	ID	yes	unique identifier (“art”+integer)
	name	CDATA	yes	article name
<i>(N.B.: The article is stored as the character data content of the article element. Links to archaeological, textual, and linguistic information can be interspersed where needed using the ad-ref, td-ref, ld-ref, vr-ref, sa-ref, and br-ref elements.)</i>				
article-ref	ref	IDREF	yes	points to an article element

<i>Element Type</i>	<i>Attribute Name</i>	<i>Attribute Type</i>	<i>Required?</i>	<i>Comments</i>
bibliography	name	CDATA	yes	<i>name of set of bibliographic citations</i>
biblio	id	ID	yes	<i>unique identifier ("bib"+integer)</i>
	name	CDATA	yes	<i>abbreviated name of bibliographic citation</i> <i>(N.B.: The content of the bibliographic citation is stored as the character data content of the biblio element.)</i>
biblio-ref	ref	IDREF	yes	<i>points to a biblio element</i>