

BAMDES IN A NUTSHELL

There are two types of metadata: external (common to every resource or tool) and internal (specific to each resource type).

External metadata (Administrative metadata)

These metadata are the basic administrative metadata necessary to identify the resource

Attribute	Type	Value description/Explanation
resourceTitle	string	Name of the resource
persistentIdentifier	string	Specification of a persistent identifier that refers to the resource or tool/service this metadata information describes
url	URI	URL address where the resource can be found
organization	string	Name of the institution who owns the resource/tool
contactEmail	string	Name and contact email of the person responsible of the resource/tool. It shall have this format: "Name Surname namesurname@email.com "

Internal metadata

These metadata are the metadata necessary to describe the main features of the resource/tool. Each resource/tool has a different metadata set according to its intrinsic characteristics.

Lexical Resources

Attribute	Type	Value description/Explanation
language	Closed vocabulary	Select the language(s) contained in the resource.
lexiconType	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - Monolingual - Bilingual - Multilingual - Computational Lexicon - Machine Readable Dictionary - Multimedia Dictionary - Other
size	Size unit: string Size: string	Here you shall specify the size of your lexicon as well as the size unit you are using (e.g. number of entries: 1652)
coverageType	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - General language - Domain specific
domain	Open vocabulary	Specify the domains covered by your lexicon (e.g. medicine, history, mathematics, etc.)
annotationFormat	Open vocabulary	Specify the format in which your lexicon is encoded (e.g. LMF, PAROLE, tabular, etc.)
informationContained	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - definition - equivalents - neologisms - example of use - collocations - selectional restrictions - phonology - morphology - syntax - semantics - phonetics - synonyms - antonyms - etymology - related entry/entries - context - source - other

Written Corpora

Attribute	Type	Value description/Explanation
corpusType	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - Monolingual - Bilingual - Multilingual - Aligned_document - Aligned_paragraph - Aligned_sentence - Aligned_word - Parallel - Comparable - Transcribed - Read - Spontaneous - Dialogue - Monologue
language	Closed vocabulary	Select the language(s) contained in the resource.
size	Size unit: string Size: string	Here you shall specify the size of your corpus as well as the size unit you are using (e.g. number of tokens: 1652)
coverageType	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - General language - Domain specific
domain	Open vocabulary	Specify the domains covered by your corpus (e.g. medicine, history, mathematics, etc.)
annotationFormat	Open vocabulary	Specify the format in which your corpus is encoded (e.g. MAF, EAGLES, etc.)
annotation	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - Full - Partial - Null
annotationLevelType	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - syntax - morphology - semantics - morphosyntax - phonology - pragmatics - lemmatization - stemming - POS tagging - Below-POS tagging - Shallow parsing - NERC (Named Entity Recognition and Classification) - SRL (Semantic Role Labeling) - Sense annotation - Time annotation - Event annotation - Topic annotation - Prosody - Gesture - Location - Age - Facial expression - Object annotation - Image categorization - Cross-media annotation - Speaker annotation - Text recognition from images and videos - Acoustic events annotation - Movement annotation - Other
tagset	String	Please indicate the tagset used in the annotation of your corpus

Oral Corpora / Multimedia corpora

Attribute	Type	Value description/Explanation
corpusType	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - Monolingual - Bilingual - Multilingual - Aligned_document - Aligned_paragraph - Aligned_sentence - Aligned_word - Parallel - Comparable - Transcribed - Read - Spontaneous - Dialogue - Monologue
language	Closed vocabulary	Select the language(s) contained in the resource.
size	Size unit: string Size: string	Here you shall specify the size of your corpus as well as the size unit you are using (e.g. number of minutes: 1652)
domain	Open vocabulary	Specify the domains covered by your corpus (e.g. medicine, history, mathematics, etc.)
annotationFormat	Open vocabulary	Specify the format in which your corpus is encoded (e.g. mp3, mpeg4, etc.)
transcription	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - Full - Partial - Null
annotation	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - Full - Partial - Null
annotationLevelType	Closed vocabulary	Select the appropriate value(s): <ul style="list-style-type: none"> - syntax - morphology - semantics - morphosyntax - phonology - pragmatics - lemmatization - stemming - POS tagging - Below-POS tagging - Shallow parsing - NERC (Named Entity Recognition and Classification) - SRL (Semantic Role Labeling) - Sense annotation - Time annotation - Event annotation - Topic annotation - Prosody - Gesture - Location - Age - Facial expression - Object annotation - Image categorization - Cross-media annotation - Speaker annotation - Text recognition from images and videos - Acoustic events annotation - Movement annotation - Other
tagset	String	Please indicate the tagset used in the annotation of your corpus

Tools

Attribute	Type	Value description/Explanation
language	Closed vocabulary	Select the language(s) contained in the resource.
runningEnvironment	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - GUI - Command line - Web service
languageDependency	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - Yes - No
operatingSystem	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - Windows - Linux - Mac-OS
task	Closed vocabulary	Select the appropriate value: <ul style="list-style-type: none"> - named entity recognition - language guessing - clustering - tokenization - format conversion - OCR - Optical Character Recognition - Spell checking - stemming/lemmatization - chunking/segmentation - analysis - morphological tagging - morphosyntactic tagging - syntactic tagging - semantic tagging - alignment - keyword extraction - lexicon/terminology extraction - statistics analysis - Word Sense Disambiguation - terminology management - summarization - text mining - machine translation - speech recognition - text to speech - querying - corpus workbench - management - indexing - knowledge representation - corpus processing - other