|  |  |
| --- | --- |
| Allowed Characters | This data category is used to specify the characters that are permitted in a given piece of content |
| allowedCharacters | Contains the regular expression indicating characters that are permitted in the annotated content. |
| Storage Size | This data category is used to specify the maximum storage size of a given content |
| storageSize | Contains the maximum number of bytes the text of the annotated content is allowed in storage. |
| storageEncoding | Contains the name of the character encoding used to calculate the number of bytes of the text in the annotated content. This may be used in calculating if the text in the annotated content can be accommodated a specified maximum storage size. The name MUST be one of the names or aliases listed in the IANA Character Sets registry. The default value is the string "UTF-8". |
| lineBreakType | Value indicates what type of line breaks the storage uses for the text in the annotated content. This may be used in calculating if the text in the annotated content can be accommodated a specified maximum storage size. The possible values are: "cr" for CARRIAGE RETURN (U+000D), "lf" for LINE FEED (U+000A), or "crlf" for CARRIAGE RETURN (U+000D) followed by LINE FEED (U+000A). The default value is "lf". |
| Domain | This data category is used to identify the topic or subject of content. Such information allows for more relevant linguistic choices during various processes. |
| domains | Used to identify the topics or subjects of content. Value is a space separated list of strings each representing a topics or subject. Strings containing spaces it must delimited by escaped quotation marks, that is pairs of APOSTROPHE (U+0027) or QUOTATION MARK (U+0022). Such information allows for more relevant linguistic choices during various processes. |
| Localisation Note | This Data Category is used to communicate notes to localizers about a particular item of content. |
| locNote | Used to communicate information about the content to localizers. |
| locNoteRef | An IRI referring to the location of a note used to communicate information about the content to localizers. |
| locNoteType | Has value ‘alert’ when a localisation note for the annotated content contains information that the translator has to read before translating a piece of text. Example: an instruction to the translator to leave parts of the text in the source language, or has value ‘description’ when the note provides useful background information that the translator will refer to only if they wish. Example: a clarification of ambiguity in the source text. |
| Localization Quality Issue | This data category is used to express information related to localization quality assessment tasks. Such tasks can be conducted on the translation of some source content (such as a text or an image) into a target language or on the source content itself where its quality may impact on the localization process. |
| LocQualityIssue | An objects used to annotate content with details of a localization quality issue and how this issue should be processed. |
| locQualityIssueComment | The human-readable description of the annotated localization quality issue. |
| locQualityIssueEnabled | A flag indicating whether the annotated localization quality issue is enabled or not. It has a value of yes or no, with the default value being yes. |
| locQualityIssueProfileRef | An IRI referencing a description of the quality assessment model or a specific profile (customization/instantiation) of a model, where relevant, used for recording and processing the annotated localization quality issue. |
| locQualityIssueSeverity | An indicator of the seriousness of the annotated localization quality issue. The seriousness depends on the Quality Model that is being applied. The Quality Model should be made explicit via the Profile Reference. The value is a rational number in the interval 0 to 100 (inclusive). The higher values represent greater severity. |
| locQualityIssueType | A classifier that assigns the annotated localization quality issue into a category (for example to differentiate spelling errors from grammar errors). The value is one of categories defined at: http://www.w3.org/TR/its20/#lqissue-typevalues |
| locQualityIssuesRef | An IRI referencing a localization quality issue object relevant to the annotated content. |
| Localization Quality Rating | This data category is used to express an overall measurement of the localization quality of a document or an item in a document. |
| locQualityRatingProfileRef | An IRI pointing to the reference document describing the quality assessment model used for the scoring the localization quality of the annotated content. |
| locQualityRatingScore | Used express an overall measurement of the localization quality of the annotated content. Its value is a rational number in the interval 0 to 100 (inclusive). The higher values represent better quality. |
| locQualityRatingScoreThreshold | Use to indicate the lowest score that constitutes a passing score in the profile used when rating the localization quality of the annotated content. Its value is a rational number in the interval 0 to 100 (inclusive). |
| locQualityRatingVote | Used express the result of a vote on the localization quality of the annotated content. Its value is a signed integer with higher values indicating a better vote. |
| locQualityRatingVoteThreshold | Used to indicate the lowest value that constitutes a passing vote in the profile used when voting on the localization quality of the annotated content. Its value is a signed integer. |
| Locale Filter | This data category specifies that a node is only applicable to certain locales. |
| localeFilterList | Used to indicate the locales for which it is applicable to localize the annotated content. It values is a comma separate list of extended language ranges conforming to BCP47 (include the ‘\*’ character for wildcards), or an empty string. |
| localeFilterType | Has value ‘include’ if the annotated content is applicable to the locales specified by a localeFilterList property of the annotated content. Has value ‘exclude’ if the annotated content is not applicable to the locales specified by a localeFilterList property of the annotated content. |
| MT Confidence | This data category is used to communicate the confidence score from a machine translation engine for the accuracy of a translation it has provided. |
| mtConfidence | Used to communicate the confidence score from a machine translation engine for the accuracy of an automated translation that resulted in the annotated content. Its value is a rational number in the interval 0 to 1 (inclusive). |
| mtConfidenceAnnotatorsRef | An IRI that references the processor that generated the mtConfidence annotation for the annotated content. |
| Provenance | This data category is used to communicate the identity of agents that have been involved in the translation of the content or the revision of the translated content. This allows translation and translation revision consumers, such as post-editors, translation quality reviewers, or localization workflow managers, to assess how the performance of these agents may impact the quality of the translation. Translation and translation revision agents can be identified as a person, a piece of software or an organization that has been involved in providing a translation that resulted in the selected content. |
| ProvRecord | An object used to annotate content with details of it provenance related to localization. Translation and translation revision agents can be identified as a person, a piece of software (tool) or an organization that has been involved in providing a translation that resulted in the annotated content. |
| org | Used to identify an organisation involved in the translation that resulted in the annotated content. |
| orgRef | An IRI referencing an organisation involved in the translation that resulted in the annotated content. |
| person | Used to identify a person involved in the translation that resulted in the annotated content. |
| personRef | An IRI referencing a person involved in the translation that resulted in the annotated content. |
| provRef | An IRI referencing external provenance information. |
| provenanceRecordsRef | An IRI referencing a ProvRecord object instance. |
| revOrg | Used to identify an organisation involved in the translation revision that resulted in the annotated content. |
| revOrgRef | An IRI referencing an organisation involved in the translation revision that resulted in the annotated content. |
| revPerson | Used to identify a person involved in the translation revision that resulted in the annotated content. |
| revPersonRef | An IRI referencing a person involved in the translation revision that resulted in the annotated content. |
| revTool | Used to identify a software tool involved in the translation revision of the annotated content. |
| revToolRef | An IRI referencing a software tool involved in the translation revision of the annotated content. |
| tool | Used to identify a software tool involved in the translation that resulted in the annotated content. |
| toolRef | An IRI referencing a software tool involved in the translation that resulted in the annotated content. |
| Text Analysis | This data category is used to annotate content with lexical or conceptual information for the purpose of contextual disambiguation. This information can be provided by so-called text analysis software agents such as named entity recognizers, lexical concept disambiguators, etc., and is represented by either string valued or IRI references to possible resource descriptions. |
| taClassRef | An IRI referencing the entity type or concept class information associated through text analysis of the annotated content. |
| taConfidence | The confidence of the text analysis processing agent in its computation of the text analysis annotations it provided of the annotated content. |
| taIdent | An identifier referencing the entity or concept information associated with the annotated content using automated text analysis. The identifier should be unique within the collection resource identified in the taSource attribute of the annotated content. |
| taIdentRef | An IRI referencing the entity or concept information associated with the annotated content using automated text analysis. |
| taSource | Identifies the collection resource from which the identifier that the taIdent property gives for the annotated content is sourced |
| termConfidenceAnnotatorsRef | An IRI that references the processor that generated the taConfidence annotation for the annotated content. |
| Terminology | Data category is used to mark terms and optionally associate them with information, such as definitions. This helps to increase consistency across different parts of the documentation. It is also helpful for translation. |
| term | The its-term attribute is used to mark terms so they can be recognized during translation process. This helps to increase consistency of translation across different parts of the documentation. It is also helpful for translation. |
| termConfidence | Used to indicate the confidence of the agents producing the annotation that the annotated unit is a term or not. Its value is a rational number in the interval 0 to 1 (inclusive). 1 represents the highest level of confidence. |
| termInfoRef | An IRI that references a resource with information about the annotated term. |
| termConfidenceAnnotatorsRef | An IRI that references the processor that generated the termConfidence annotation for the annotated content. |
| Translate | This data category expresses information about whether the content of an element or attribute is intended for translation or not. |
| translate | Used express information about whether the content of an element or attribute is intended for translation or not. Its value is either "yes" (translatable) or "no" (not translatable). |
| Element Within Text | This data category reveals if and how an element affects the way text content behaves from a linguistic viewpoint. This information is for example relevant to provide basic text segmentation hints for tools such as translation memory systems. |
| within-text | Used to indicate the relationship between some marked up text element and the surrounding text. It has values: “yes” if the text element is part of the flow and content of the surrounding text, e.g. text with emphasis markup; “nested” if the text is part of the flow of the surrounding text but the content is not, e.g. text annotated as a footnote; and marked up text that splits the flow from the surrounding text, e.g. an individual bullet within a paragraph. |