



## SHACHI メタデータセット

## 目 次

<b>1. DCMI Metadata Terms</b> .....	1
1.1. DCMI とは.....	1
1.2. 概要.....	1
1.3. メタデータ用語の仕様の定義.....	1
1.3.1. 各メタデータ用語の仕様の定義.....	1
<b>2. OLAC Metadata</b> .....	2
2.1. OLAC とは.....	2
2.2. 概要.....	2
2.3. OLAC メタデータ (OLAC Standards) .....	2
<b>3. SHACHI メタデータセット</b> .....	2
3.1. Term Name: <b>contributor</b> .....	3
3.1.1. OLAC: Participant Roles [olac:role].....	3
3.1.1.1. annotator.....	3
3.1.1.2. author.....	3
3.1.1.3. compiler.....	3
3.1.1.4. consultant.....	3
3.1.1.5. data_inputter.....	4
3.1.1.6. depositor.....	4
3.1.1.7. developer .....	4
3.1.1.8. editor .....	4
3.1.1.9. illustrator.....	4
3.1.1.10. interpreter.....	4
3.1.1.11. interviewer.....	4
3.1.1.12. participant.....	4
3.1.1.13. performer .....	4
3.1.1.14. photographer.....	5
3.1.1.15. recorder .....	5
3.1.1.16. researcher .....	5
3.1.1.17. research_participant.....	5

3.1.1.18.	responder.....	5
3.1.1.19.	signer.....	5
3.1.1.20.	singer.....	5
3.1.1.21.	speaker.....	5
3.1.1.22.	sponsor.....	5
3.1.1.23.	transcriber.....	5
3.1.1.24.	translator.....	6
3.1.2.	SHACHI: Attribute of contributor.....	6
3.1.2.1.	SHACHI: mother-tongue.....	6
3.1.2.2.	SHACHI: intonation.....	6
3.1.2.3.	SHACHI: level.....	6
3.1.2.4.	SHACHI: age.....	6
3.1.2.5.	SHACHI: gender.....	6
3.2.	Term Name: <b>coverage</b> .....	6
3.2.1.	DC refinement: Temporal.....	7
3.2.2.	DC refinement: Spacial.....	7
3.3.	Term Name: <b>creator</b> .....	7
3.4.	Term Name: <b>date</b> .....	8
3.4.1.	DC refinement: Created.....	9
3.4.2.	DC refinement: Issued.....	9
3.5.	Term Name: <b>description</b> .....	9
3.5.1.	SHACHI: Price.....	10
3.6.	Term Name: <b>format</b> .....	10
3.6.1.	DC refinement: Extent.....	10
3.6.2.	DC refinement: Medium.....	10
3.6.3.	SHACHI:functionality.....	10
3.6.4.	SHACHI:markup.....	10
3.6.5.	SHACHI:encoding.....	11
3.7.	Term Name: <b>identifier</b> .....	11
3.8.	Term Name: <b>language</b> .....	11
3.9.	Term Name: <b>publisher</b> .....	12
3.10.	Term Name: <b>relation</b> .....	12
3.10.1.	isVersionOf.....	13
3.10.2.	hasVersion.....	13
3.10.3.	isReplacedBy.....	13
3.10.4.	Replaces.....	13

3.10.5.	isRequiredBy .....	13
3.10.6.	Requires .....	14
3.10.7.	isPartOf .....	14
3.10.8.	hasPart .....	14
3.10.9.	isReferencedBy .....	14
3.10.10.	References.....	14
3.10.11.	isFormatOf.....	14
3.10.12.	hasFormat .....	15
3.10.13.	conformsTo.....	15
3.10.14.	SHACHI: Utilization.....	17
3.11.	Term Name: <b>rights</b> .....	17
3.12.	Term Name: <b>source</b> .....	17
3.13.	Term Name: <b>subject</b> .....	18
3.13.1.	OLAC: Linguistic Subject Vocabulary .....	18
3.13.1.1.	anthropological_linguistics.....	18
3.13.1.2.	applied_linguistics .....	18
3.13.1.3.	cognitive_science:.....	19
3.13.1.4.	computational_linguistics .....	19
3.13.1.5.	discourse_analysis .....	19
3.13.1.6.	forensic_linguistics .....	19
3.13.1.7.	general_linguistics.....	19
3.13.1.8.	historical_linguistics.....	19
3.13.1.9.	history_of_linguistics.....	19
3.13.1.10.	language_acquisition .....	20
3.13.1.11.	language_documentation .....	20
3.13.1.12.	lexicography .....	20
3.13.1.13.	linguistics_and_literature.....	20
3.13.1.14.	linguistic_theories.....	20
3.13.1.15.	mathematical_linguistics.....	20
3.13.1.16.	morphology .....	20
3.13.1.17.	neurolinguistics.....	21
3.13.1.18.	philosophy_of_language.....	21
3.13.1.19.	phonetics.....	21
3.13.1.20.	phonology.....	21
3.13.1.21.	pragmatics .....	21
3.13.1.22.	psycholinguistics .....	21

3.13.1.23.	semantics .....	21
3.13.1.24.	sociolinguistics .....	22
3.13.1.25.	syntax .....	22
3.13.1.26.	text_and_corpus_linguistics .....	22
3.13.1.27.	translating_and_interpreting .....	22
3.13.1.28.	typology.....	22
3.13.1.29.	writing_systems .....	22
3.13.2.	OLAC-Language extension.....	23
3.13.3.	SHACHI: mono_multi_lingual.....	23
3.13.3.1.	monolingual <SHACHI value> .....	23
3.13.3.2.	multilingual <SHACHI value> .....	23
3.13.4.	SHACHI: Resource Subject.....	23
3.13.4.1.	corpus <SHACHI value> .....	23
3.13.4.2.	dictionary <SHACHI value> .....	23
3.13.4.3.	thesaurus <SHACHI value> .....	23
3.13.4.4.	glossary <SHACHI value> .....	23
3.14.	Term Name: <b>title</b> .....	24
3.14.1.	Term Name: alternative.....	24
3.15.	Term Name: <b>type</b> .....	24
3.15.1.	The DCMI Type Vocabulary.....	25
3.15.1.1.	Image.....	25
3.15.1.2.	Sound.....	25
3.15.1.3.	Text.....	25
3.15.2.	Linguistic Data Types .....	26
3.15.2.1.	lexicon .....	26
3.15.2.2.	primary_text .....	26
3.15.2.3.	language_description.....	26
3.15.3.	Discourse Types.....	27
3.15.3.1.	drama .....	27
3.15.3.2.	formulaic_discourse .....	27
3.15.3.3.	interactive_discourse .....	27
3.15.3.4.	language_play .....	27
3.15.3.5.	oratory .....	27
3.15.3.6.	narrative .....	27
3.15.3.7.	procedural_discourse .....	27
3.15.3.8.	report.....	28

3.15.3.9.	singing.....	28
3.15.3.10.	unintelligible_speech .....	28
3.15.4.	SHACHI: Purpose .....	28
3.15.4.1.	lexicography <SHACHI value> .....	28
3.15.4.2.	analysis <SHACHI value> .....	28
3.15.4.3.	developing_technologies .....	28
3.15.4.4.	education .....	28
3.15.5.	SHACHI: Style .....	28
3.15.5.1.	speech <SHACHI value>.....	28
3.15.5.2.	written <SHACHI value>.....	28
3.15.6.	SHACHI: Form.....	29
3.15.6.1.	fixed <SHACHI value>.....	29
3.15.6.2.	unfixed <SHACHI value> .....	29
3.15.7.	SHACHI: Sentence.....	29
3.15.7.1.	short <SHACHI value> .....	29
3.15.7.2.	long <SHACHI value>.....	29
3.15.7.3.	mixed <SHACHI value>.....	29
3.15.8.	SHACHI: Annotation .....	29
3.15.8.1.	annotated <SHACHI value> .....	29
3.15.8.2.	plain <SHACHI value> .....	29
3.15.9.	SHACHI: Annotaion_sample .....	30
3.15.10.	SHACHI: Sample .....	30

## 1. DCMI Metadata Terms

ダブリン・コア メタデータ用語

<http://dublincore.org/documents/dcmi-terms/>

発効日: 2006-12-18(最新版)

### 1.1. DCMI とは

The Dublin Core Metadata Initiative (DCMI) は、相互操作性のあるメタデータスタンダードの広い適用を促進するために設立された組織で、また、資源を記述説明する目的に特化したメタデータのボキャブラリを開発している。こうしたメタデータによって、現在より高い性能を持つ情報発見システムを実現する。DCMIに関する詳細は：<http://dublincore.org/about/>

### 1.2. 概要

Dublin Core Metadata Initiative が管理しているメタデータ用語の正式な仕様。

- elements: エレメント The Dublin Core Metadata Element Set (DCMES)
  - element refinements: エレメントリファインメント (エレメントを詳細にするために使う)
  - encoding schemes: エンコーディング・スキーマ
  - vocabulary terms (the DCMI Type Vocabulary): ボキャブラリー用語 (DCMIType 言語)
- ※ 15 個のエレメントは NISO<sup>1</sup>スタンダード Z39.85-2001 と ISO<sup>2</sup>スタンダード 15836-2003

例①: element (エレメント)

```
<coverage>United States</coverage>
```

- < >で囲まれた文字列: メタデータ用語。値が「coverage」であるという意味づけを始める。
- </ >スラッシュが前についている< >内の文字列: メタデータ言語で、意味づけを終える。
- United States: 値 (Value) ここに書かれた内容が人間の読めるインターフェイスに表示される。
- 登録者は値 (United States) のみを入力しているが coverage として識別可能に (qualify) できる。

例②: element refinements, vocabulary terms

```
<dcterms:created xsi:type="dcterms:W3C-DTF">2002-11-28</dcterms:created>
```

- dc (ダブリンコア) の リファインメント である created (Date のうち作成日であることを意味する)
- xsi:type は type の値として国際標準化機構の W3C-DTF を指定している。
- 値である「2002-11-28」は W3C-DTF に準拠した書き方である必要がある (Date 参照)

### 1.3. メタデータ用語の仕様の定義

各メタデータ用語は、以下の特性説明用セットに従って仕様を定義、また用語に付加的に説明を加える。

#### 1.3.1. 各メタデータ用語の仕様の定義

<sup>1</sup> 米国情報標準化機構 (National Information Standards Organization : NISO)

<sup>2</sup> 国際標準化機構 (International Organization for Standardization : ISO) 工業標準の策定を目的とする国際機関で、各国の標準化機関の連合体。

- **Name:** メタデータ用語の名前。その用語に付与された唯一的な特徴
- **URI:** “Uniform Resource Identifier” 用語を唯一的に特定するのに使用
- **Label:** 各用語に付与された、人間に読めるラベル
- **Definition:** 定義。各用語のコンセプトや主な性格を現す記述
- **Type of Term:** 用語の種類。DCMI 文法原則で説明しているエレメント、エンコーディング・スキーマなど
- **Status:** ステータス。DCMI の使用用法委員会で決めている
- **Date issued:** 用語が最初に申告された日時
- **Comment:** コメント。メタデータ用語適用に当たっての追加情報

## 2. OLAC Metadata

### 2.1. OLAC とは

OLAC は the Open Language Archives Community の頭文字。1) 言語資源のデジタルアーカイブの最良の実施モデルのための合意を発展させること、2) その資源へのアクセスや資源の置き場所の整備において相互操作性の高いリポジトリやサービスのネットワークを開発すること、を目指して言語資源のバーチャルなライブラリを作っている団体や個人に対する国際的な協力団体である。

### 2.2. 概要

OLAC メタデータセットは、ダブリンコア (DC) メタデータセットに基づき、DC スタンドで定義される全ての 15 のエレメントを使っている。より詳細に資源を記述するために、OLAC は、DC 推奨言語 (DC Recommendation) に従い、エレメントリファインメントあるいはエレメントスキーマを使用してエレメントをより限定的に識別していく。DC 認定のすべてのエレメント、リファインメント、スキーマは DCMT を参照。

DC に推奨される識別子 (qualifier) は広範な資源に適用可能だが、それでも言語資源が要求する記述方法は一般的なスタンダードではまかなえない。このニーズに応えるために、OLAC では OLAC コミュニティ独自の qualifier を開発した。OLAC コミュニティはすでに広くこれらのうちいくつかを使っている。

### 2.3. OLAC メタデータ (OLAC Standards)

- DCMI の 15 のエレメント
- DCMI のリファインメント
- DCMI のスキーマ (登録作業者に直接関係はない)
- OLAC Extension

## 3. SHACHI メタデータセット

OLAC のメタデータは The Dublin Core Metadata Element Set (ダブリンコア メタデータエレメントセット, DCMES) 15 の基本とそれに対するリファインメントに準拠しているため SHACHI もこれに準拠していくが、SHACHI 独自の拡張メタデータ用語もある。SHACHI メタデータ全体像は別紙。

登録上の諸注意:

- 自由記入欄と選択欄がある。選択項目については各メタデータ用語を参照。
- 各メタデータ言語は一つの資源の記述のために複数回使用できるため、本マニュアル内で「別個の項を立てる」という場合、SHACHI では縦棒区切り”|”で改行すること。ただし”+、-”のボタンを操作して項を増やせる部分もあるので、その場合は新規の入力欄に別々に記入。
- " mother-tongue ", "Intonation", "Level", "Age", " Gender", “Style” "Form", "Sentence"は、辞書・シソーラス・語彙リストの登録時には、入力が必要なし。

### 3.1. Term Name: **contributor**

URI: <http://purl.org/dc/elements/1.1/contributor>

Label: Contributor 貢献者

Definition: 資源内容の作成に貢献した責任者のデータ

Comment: Contributor の例として、人、組織、サービスがある。たいていは Contributor の名前。

Type of Term: element

Status: recommended 推奨

Date Issued: 1999-07-02

#### 3.1.1. OLAC: Participant Roles [olac:role]

<http://www.language-archives.org/REC/role.html>

##### 3.1.1.1. annotator

Definition: 当該資源あるいは関係資源のアノテーション作成者

##### 3.1.1.2. author

Definition: 資源の原作者

##### 3.1.1.3. compiler

Definition: 資源の下位資源（部分）の収集者

Description: 複数の資源からひとつの資源を作る人。例として、短編集、別々にアーカイブされている資源でひとつのコーパスを作る人など。

Example: 短編集、CD のソング集などの収集者。複数の言語や、ある主題についての録音のコーパスの収集者。

##### 3.1.1.4. consultant

Definition: 資源の作成の専門知識的提供者

Example: 分析、研究対象となる言語の話し手で被験者。その言語を記述、分析するときフィールド言語学者が使う。被験者は彼らの土地の言葉における知識を資源に提供するが、彼らの発話、記号、正書法は直接資源には現れない。Informant（インフォーマント）と呼ぶ場合もある。



#### 3.1.1.5. data\_inputter

**Definition:**入力、タイピング修正、および資源が含むデータの構造化の担当者

#### 3.1.1.6. depositor

**Definition:**アーカイブへの資源保管の担当者

#### 3.1.1.7. developer

**Definition:**資源を成り立たせる方法論やその道具、あるいは資源作成に使われた方法論や道具の開発者

**Example:** ソフトウェアプログラマー, デザイナー, あるいは分析者。インフォーマントへの質問作成者や研究課題設定者。

#### 3.1.1.8. editor

**Definition:**資源の査読、修正、および試用の担当者

**Example:** プルーフリーダー, デバッガー (バグ発見), 試験者など編集的な性質を持つ役割。Compiler と重複する場合もある。

#### 3.1.1.9. illustrator:

**Definition:**資源の絵やイラストの提供者

#### 3.1.1.10.interpreter

**Definition:**現場での通訳、資源に記録された談話の解説者

**Comment:** 現場で通訳する人は **Interpreter**。談話の種類によってはその談話などをその場で解釈して説明する人。

#### 3.1.1.11.interviewer

**Definition:**資源の部分を構成するインタビューを実施した人

#### 3.1.1.12.participant

**Definition:**資源作成の間に参加していたが資源内容への貢献は多くない人

**Comment:** マイナーな参加者。観衆や周辺の参加者。資源に実際に参加していない人, オンライン議会の傍聴者や, 資源作成の時にそこに居た, など (ある市の集会での市長など)。

#### 3.1.1.13.performer

**Definition:**記録、録画、書き下された資源の一部を実際に遂行した人

**Comment:** 創造的な参加者, 'speaker', 'signer', or 'singer'.のように, より詳細に分類することができないときに使う。

**Example:**式典を行う人, 演説または説教をする人;例えば儀式挨拶または侮辱の言葉などで, 伝統的な言い方を再現しているネイティブスピーカー。俳優または音楽家。

#### 3.1.1.14. photographer

Definition: 資源に含まれる構成資源の撮影、録画をした人

#### 3.1.1.15. recorder

Definition: 資源作成のために使われる録音機械の操作をした人

#### 3.1.1.16. researcher

Definition: 資源が研究のために作られた、あるいは研究調査の結果あるいは中間結果として資源ができた場合のその研究者

#### 3.1.1.17. research\_participant

Definition: 研究の調査対象者あるいは、アンケート回答者で、その結果が資源の基盤となるような場合の対象者や回答者

#### 3.1.1.18. responder

Definition: 談話事象においてなんらかの対話者となった人

Comment: 声の参加はしている（あるいはその文書化されたものにも参加している）が、あいづちのみの聞き手。

#### 3.1.1.19. signer

Definition: 録音や録画、記録された資源の書き下し文書の中で、手振り身振り手話などのサインを使った主な人

Comment: 録画された資源でジェスチャーをしている人（録画資源の文書化資源でもいい）。

#### 3.1.1.20. singer

Definition: 録音、録画、あるいは記録資源の書き下し文などで成る資源において独唱あるいは合唱の一員として歌った人

#### 3.1.1.21. speaker

Definition: 録音、録画、あるいは記録資源の書き下し文などで成る資源における主な話し手

Comment: 録音、録画された資源において優先的に話をしている人（文書化されていてもいい）

Example: 録音された会話、会話引き出しによるセッションまたは非公式のナレーションの参加者。時おりしか参加しない聞き役は **Responder** に分類。

#### 3.1.1.22. sponsor

Definition: 資源作成の資金援助者

#### 3.1.1.23. transcriber

Definition: 当該あるいは関連資源の書き下し文作成者

### 3.1.1.24. translator

Definition: 当該あるいは関連資源の翻訳者

### 3.1.2. SHACHI: Attribute of contributor

旧 Sender 情報, Contributor のうちで **speaker**, **author** のような資源の主な話し手聞き手の情報。

#### 3.1.2.1. SHACHI: mother-tongue

資源の話し手や書き手が **native** か **non-native** を選択。

#### 3.1.2.2. SHACHI: intonation

地理的な方言の状況。話し手あるいは書き手が標準あるいは方言を使っているかどうか、**standard** か **dialect** で選択。

#### 3.1.2.3. SHACHI: level

Definition: 話し手や書き手が専門的言語的な訓練（話す、書くなど）を受けたことがあるかどうか、**professional** か **amateur** で選択。

#### 3.1.2.4. SHACHI: age

Description: 話し手や書き手の年齢。 **Adult** か **child** を選択。大勢いる場合、この人々が属するすべての年齢層の間の比率。

#### 3.1.2.5. SHACHI: gender

Description: 話しか書き手の性別。 **Male**, **female**, **male & female** のうちから選択。大勢いる場合は男女比。

## 3.2. Term Name: **coverage**

URI: <http://purl.org/dc/elements/1.1/coverage>

Label: Coverage カバー範囲

Definition: 資源の空間的（どこ）及び時間的（いつ）に関すること。資源の空間的な適用性、資源の管轄団体や関連する地名。

Comment: 空間的な主題とは場所名や地理的座標で特定できる場所など。時間的な期間とは期間名、日付、日付の範囲など。管轄地域とは、資源が当てはまる管轄地区名や地理的な場所のこと。 **Thesaurus of Geographic Names [TGN]** のような統制言語の使用が望ましい。<sup>3</sup>

References: [TGN] <http://www.getty.edu/research/tools/vocabulary/tgn/index.html>

Type of Term: element

---

<sup>3</sup> SHACHI では統制言語を選択欄で提示するので上記 URL の参照は不要。

Status: recommended

Date Issued: 1999-07-02

※OLAC では、サービスプロバイダは languages をそれが話されている countries にマッピングしてあるデータベースを持っている。OLAC-Language 参照

Coverage を、この情報を繰り返さず使用しないように：サービスプロバイダは当該の国で話される言語を言語データベースを参照して検索し易くするほうがいい。

Coverage はその場合地理データに使用、ただしそれは当該言語が話される地域が広すぎるケースや、資源が特定地域や地理的な管轄地域をカバーするとき、あるいは言語資源が複数言語地域に当たる場合にのみ。

#### Examples

- ・ インドの英語についての資源:

```
<subject.language code="en"/>
```

```
<coverage>India</coverage>
```

- ・ ガタルカナル島で話される言語についての資源:

```
<coverage scheme="TGN">Guadalcanal (island)</coverage>
```

- ・ 19 世紀に使われた言語についての資源: <coverage>19th century</coverage>

#### 3.2.1. DC refinement: Temporal

Definition: 資源の知的内容の時間的な性質

Comment: 言語資源データの時代. yyyy-mm-dd で記載. 講演など、現代に行われているものが収録されている場合は、Date.created の年代と同じになります.

#### 3.2.2. DC refinement: Spatial

Definition: 資源の知的内容の空間的な性質

Comment: 空間的な主題とは場所名や地理的座標で特定できる場所など。管轄地域とは、資源が当てはまる管轄地区名や地理的な場所のこと。

推奨: Thesaurus of Geographic Names[TGN] 統制言語として使用。

<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>

#### 3.3. Term Name: creator

URI: <http://purl.org/dc/elements/1.1/creator>

Label: Creator 作成者

Definition: 資源内容の作成の主要責任者

Comment: Creator の例として、人、団体、サービスがある。たいてい Creator の名前を記述する。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※OLAC では Creator の名前を入れるとき、名前はインデックスでソートできるように準備された形式推奨。人の名前は苗字先頭の形式推奨。団体名の先頭の a/the など冠詞は省略すること。

一人以上の Creator がいるときは別々のエレメントを立てる。

[OLAC-Role]使用可能。※SHACHI では選択欄から選択できる。

Examples:

個人の著者: <creator>**Bloomfield, Leonard**</creator>

機関が著作している: <creator>**Linguistic Society of America**</creator>

編集者: <creator refine="editor">**Sapir, Edward**</creator>

# Role (それぞれの定義は別紙参照)

* annotator	* illustrator	* research_participant
* author	* interpreter	* responder
* compiler	* interviewer	* signer
* consultant	* participant	* singer
* data_inputter	* performer	* speaker
* depositor	* photographer	* sponsor
* developer	* recorder	* transcriber
* editor	* researcher	* translator

### 3.4. Term Name: **date**

URI: <http://purl.org/dc/elements/1.1/date>

Label: Date 日付

Definition: 資源のライフサイクルにおける出来事のあった時点(作成日や公開日など)

Comment: 日付を使って様々なレベルの時間的情報を記す。エンコーディング・スキームを使うことを推奨 (ISO 8601 の W3CDTF プロファイルなど)。

References: [W3CDTF] <http://www.w3.org/TR/NOTE-datetime>

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※ SHACHI では入力 BOX に入力。YYYY-MM-DD と記述される。自由記入欄に記入する際は以下の通り。

- 10 以下の数字は 0 をつけて 2 桁にする (1999-05-01 など)
- 年号のみ YYYY
- 年号と月のみ YYYY-MM
- 正確な日付 YYYY-MM-DD
- 年号に範囲があるとき: 両方 4 桁にすること, 早い年号が前にくこと。YYYY-YYYY

※万一, 時間 (分や秒) までの記述が資源の説明記述にあたって重要要素である場合は W3CDTF 参照。

refinement もオプションで使用できる。統制言語で, 作成年月日に対して出版年月日 **修正年月日**など。

Date 用のボキャブラリは[DC-Q]で定義されている。DC のリファインメント。

※OLAC では date 属性での定義は重複しないように (ひとつの資源に出版年 2 つなど)。2 つ以上ある場合, サービスプロバイダは最も新しい日付を, 作成, 出版, 修正の日付として選択する。

Examples

出版年:<date code="1992"/>

修正年月日が October 16, 1996:<date refine="modified" code="1996-10-16"/> (SHACHI では未使用)  
1950 年ぐらいからの資源 1950:<date code="1950">circa1950</date><sup>4</sup>

#### 3.4.1. DC refinement: Created

Definition: 資源の作成日

Comment: 言語資源が制作された年月日. yyyy-mm-dd で記載. 例) 1999-2003, 2003-07-22, 2005-12.

#### 3.4.2. DC refinement: Issued

Definition: 資源の公的な公開日 (出版日など)

Comment: 言語資源が配布された年月日. yyyy-mm-dd で記載. 例) 2007, 2002-08-11

### 3.5. Term Name: **description**

URI: <http://purl.org/dc/elements/1.1/description>

Label: Description

Definition: 資源内容の説明

Comment: 概要, 目次, 写真の表示や図示, あるいは資源内容の自由記述を記入するが, これに限らず自由に記入可能

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※OLAC では, Description テキストへの規定フォーマットは無い。サービスプロバイダは全文を一段落にフォーマットする。段落分けしても無効になるので注意。

資源説明をしている URL があるとき, Description エlement をもうひとつ別に立てて, URL とだけエンコードする。"http:"で始まる Description は URL としてサービスプロバイダに解釈されるので注意。その場合ユーザーインターフェイスではリンクとして表示される。サービスプロバイダは Description 中に URL がないかどうか探する必要は無い。

#### Examples

- ・ 資源のプロセス説明:

```
<description>The CALLHOME Japanese corpus of telephone speech consists of 120 unscripted telephone conversations between native speakers of Japanese. All calls, which lasted up to 30 minutes, originated in North America and were placed to locations overseas (typically Japan). Most participants called family members or close friends. This corpus contains speech data files ONLY, along with the minimal amount of documentation needed to describe the contents and format of the speech files and the software packages needed to uncompress the speech data. </description>
```

- ・ オンライン上に存在する資源への参照:

```
<description>http://www ldc upenn edu/Catalog/LDC96S37.html</description>
```

<sup>4</sup> circa: 【前】〈ラテン語〉およそ、約、～年頃

### 3.5.1. SHACHI: Price

Definition: ある媒体でのその資源の価格情報。

Comment: 言語資源が配布されている場合、有料であれば、その価格を記載。無料であれば"フリー"と記載

### 3.6. Term Name: **format**

URI: <http://purl.org/dc/elements/1.1/format>

Label: Format フォーマット

Definition: 資源のファイル形式、物理的媒体、外寸など。

Comment: 寸法の例として、サイズや期間がある。Internet Media Types [MIME]のような統制言語の使用を推奨。

References: [MIME] <http://www.iana.org/assignments/media-types/>

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※ DC の Refinement 2つと SHACHI の拡張 3つ。

#### 3.6.1. DC refinement: Extent

Definition: 資源のサイズや期間

※ SHACHI における登録詳細は別紙参照

#### 3.6.2. DC refinement: Medium

Definition: 資源の材料や搬送媒体

※ SHACHI における登録詳細は別紙参照

#### 3.6.3. SHACHI:functionality

Name ファンクショナリティ (Format の拡張)

Definition 資源の搬送媒体にセットされているソフトウェアの情報。

Comments 言語資源に付属しているツール。ツール名は無いが、こういうツールがあるという記載がある場合は、その部分を要約。漠然としている場合は、URL を記入。

例) 構文解析ツール, 単語ソートツール, 音声視覚化ツールなど

※ OLAC では Type で特定したソフトウェア資源の機能性を記述するのに使うことが提案されていたが、SHACHI では Type では image, text, sound しか指定しないので Type としては使用しない。

#### 3.6.4. SHACHI:markup

Name: Markup マークアップ

Definition: デジタル資源に使われているマークアップスキーム

Comments: HTML SGML XML など

### 3.6.5. SHACHI:encoding

Name: Character Encoding キャラクターエンコーディング

Definition: デジタル資源のキャラクタセットの文字コード

Comments: ASCII, EUC, シフト JIS(=MS 漢字コード)など

※ SHACHI では Format を詳述するために markup と encoding を独自に使用する。OLAC で廃止された Format.markup および Format.encoding とは定義がまったく異なるので注意。SAHCHI の定義は別紙参照。

### 3.7. Term Name: **identifier**

URI: <http://purl.org/dc/elements/1.1/identifier>

Label: Identifier アイデンティファイア (識別子)

Definition: 現状範囲での資源への明白な参照

Comment: 公式の識別システム推奨 (SHACHI では URL を使用)

- URI (Uniform Resource Identifier)
- URL (Uniform Resource Locator)
- Digital Object Identifier (DOI)
- International Standard Book Number (ISBN). など。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※ OLAC では電子的にはエンコードされていない場合で、既存のアーカイブに収まっている場合 (カセットテープ, ビデオ, 本など) Identifier はローカルでの棚番号や箱番号, あるいは資源がコレクション内で探せる場所を示すなんでも, を書くことになる。

※ "http:"で始まる Identifier はサービスプロバイダによって URL と解釈され, ユーザーインターフェイスではリンク表示される。

※ OLAC では Identifier は実際の資源のある場所の URL のみに使われる。資源解説の URL は Description 項目を使う。

※ ※SHACHI では Identifier で資源解説の URL を入れている点が違うので注意。

### 3.8. Term Name: **language**

URI: <http://purl.org/dc/elements/1.1/language>

Label: Language 言語

Definition: 資源の知的内容の言語

Comment: 統制言語使用推奨, 例: RFC 3066 [RFC3066]. ※SHACHI では OLAC-Language extension に基づき選択式。

References: [RFC3066] <http://www.ietf.org/rfc/rfc3066.txt>

Type of Term: element



Status: recommended

Date Issued: 1999-07-02

- ※ OLAC では、OLAC-Language extension で指定される The ISO639-3 extension を使用。現存の言語あるいは絶滅言語として知られているすべての言語のコードである。SHACHI では選択式。
- ※ OLAC では資源が二つ以上の言語で書かれている場合、別々の Language エレメントを立ててそれぞれの言語を書く。
- ※ SHACHI (OLAC 準拠) 選択欄に適切な言語がない場合、あるいは統制言語ですでに指定した言語であるがさらなる詳細の指定が必要な場合 (例えば方言をはっきりと明記するとき、言語に別名がある場合など) 自由記入欄に記す。

### Examples

シカイアナ語についての英語の資源:

```
<language code="en"/>
```

```
<subject.language code="x-sil-SKY"/>
```

イェンバ - フレンチ辞書でジャングという言語名がより好ましい場合

```
<language code="fr"/>
```

```
<subject.language code="x-sil-BAN">Dschang</subject.language>
```

The American Heritage Dictionary でアメリカ英語についてアメリカ英語で書かれている:

```
<language code="en"/>
```

```
<subject.language code="en"/>
```

```
<coverage>United States</coverage>
```

統制言語が付与されていない言語

```
<subject.language>Ancient Sumerian</subject.language>
```

### 3.9. Term Name: **publisher**

URI: <http://purl.org/dc/elements/1.1/publisher>

Label: Publisher 公開元, 出版社

Definition: 資源内容の頒布責任者

Comment: Publisher の例として, 人, 組織, サービスがある。たいていは公開者, 出版社の名前。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※OLAC: 人の名前は苗字先頭の形式推奨。団体名の先頭の a/the など冠詞は省略すること。一つ以上の Publisher がいるときは別々のエレメントを立てる。

### 3.10. Term Name: **relation**

URI: <http://purl.org/dc/elements/1.1/relation>

Label: Relation 関係

Definition: 関連資源

Comment: 関連資源を公式の識別システムで特定することが望ましい。※SHACHI では資源名。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※DC の refinement を使用 (以下, 3.10.1 から 3.10.13 まで)

※ SHACHI 拡張ふたつ有り (3.10.14 SHACHI: Utilization)

### 3.10.1. isVersionOf

Label: Is Version Of

Term description: 当該資源[Title]が参照資源[Relation]のバージョン、版、改作である。バージョン変更とは単にフォーマットの違いではなく内容の大幅変更を指す。

Guidelines for creation of content:

内容の変更を記述するために使用。ブロードウェイの舞台の「ウェストサイド物語」と映画版の「ウェストサイド物語」は内容の解釈などに関わるので isVersionOf.

ビデオの「ウェストサイド物語」と DVD の「ウェストサイド物語」は両方とも同じ映画なので内容は同じ。この場合は isFormatOf.

### 3.10.2. hasVersion

Label: Has Version

Term description: 当該資源[Title]が参照資源[Relation]のヴァージョン、版、改作である。ヴァージョン変更とは単にフォーマットの違いではなく内容の大幅変更を指す。

### 3.10.3. isReplacedBy

Label: Is Replaced By

Term description: 当該資源[Title]は参照資源[Relation]によって置換えられている。

Guidelines for creation of content:

続々とバージョンを増やしていく場合に、そのうち最新版だけが有効資源である場合がある。

isReplacedBy と Replaces は、ユーザーが最適のバージョンに誘導されることを目的とする。この場合相互関係の明記が非常に重要。

### 3.10.4. Replaces

Label: Replaces

Term description: 当該の資源が参照資源[Relation]を置き換えている(当該資源が有効)。

### 3.10.5. isRequiredBy

Label: Is Required By

**Term description:** 当該資源が関連資源に物理的あるいは理論的に必要とされる。

**Guidelines for creation of content:**

**Requires** の方が **IsRequiredBy** よりも明らかに必要性が高い。

もちろんどちらも有用である。最も多いケースは、ソフトウェアと書類、ソフトウェアとアプリケーション、ハードウェアとソフトウェアなど。

### 3.10.6. Requires

**Label:** Requires

**Term description:** 当該資源が関連資源を必要とする。機能させるため、搬送のため、内容の連続性のためなど。

### 3.10.7. isPartOf

**Label:** Is Part Of

**Term description:** 当該資源が関連資源の一部である。(物理的あるいは論理的に)

**Guidelines for creation of content:**

この **isPartOf** と **hasPart** の関係は本質的に「親子」関係である、つまり継承の性質を持つ。1対1か、1対多のタイプの関係が多い。

### 3.10.8. hasPart

**Label:** Has Part

**Term description:** 当該資源が関連資源を含む。(物理的あるいは論理的に)

### 3.10.9. isReferencedBy

**Label:** Is Referenced By

**Term description:** 当該の資源は関連資源に参照される、引用されるなど。関連資源から逆に指摘を受ける場合も。

**Guidelines for creation of content:**

この **isReferencedBy** と **References** はユーザーの便宜をはかることができるものだが、資源の作成意図や資源のライフサイクルとは関係ない。ある記事がある資源を批評している場合や、オリジナルの発話の風刺スピーチなど。

### 3.10.10. References

**Label:** References

**Term description:** 当該の資源が関連資源を参照、引用、指摘している。

### 3.10.11. isFormatOf

**Label:** Is Format Of

**Term description:** 当該の資源は関連資源と同じ内容(知的内容)の資源だが、別フォーマットで提示されている。

**Guidelines for creation of content:**

この関係は資源と資源の関係がそのフォーマットが不定であるときに使う。Dublin Core は 1 対 1 原則なので、各資源が個々に記述されることが望ましい。

**3.10.12. hasFormat**

**Label:** Has Format

**Term description:** 当該資源が関連資源より前に存在していた。その関連資源は同じ内容(知的内容), 別フォーマット。

**3.10.13. conformsTo**

**Label:** Conforms To

**Term description:** 水準があつて資源が準拠している。

**Guidelines for creation of content:**

教育水準, アクセシビリティ基準, ユーザー補助基準, その他どのような水準, 基準でも資源に使用されるもの。

## ※ ダブリンコアにおける具体例

<http://dublincore.org/documents/usageguide/elements.shtml#relation>

レベル : Relation

この要素の説明 : 関連ソースについて記入

ガイドライン :

- 資源と資源の関係性は逆もまた成り立つ場合もあれば, 一方向的なこともある。
- もし資源 ID などを明記しない場合, 資源名は公式の名前を入力する必要がある。
- 必ず URI を使う。URI を使うときスキーマは指定されること (SHACHI では未使用)

以下は具体例

- IsPartOf.

2 巻本の小説「2つの生涯」で, そのうち一冊が「ツルゲーネフを読む」というタイトル。

Title= “ツルゲーネフを読む”

Title="Reading Turgenev" 2つの生涯

Relation="Two Lives" ツルゲーネフを読む

[Relationship described is **IsPartOf**.

Relation= “二つの生涯” IsPartOf.

この場合, A is part of B の関係性は以下。

“資源名(Title)「ツルゲーネフを読む」 is part of “2 巻本「2つの生涯」”

このように物理的に (本 2 冊のうち一冊) な全体部分関係のみではなく, 理論的に全体部分関係になつ

ていることもある。

- IsVersionOf

Title="Candle in the Wind" エルトン・ジョンのダイアナ妃追悼曲

Relation="Elton John's 1976 song Candle in the Wind" 1976年の旧版

[Relationship described is **IsVersionOf**. 登録中の資源は同じもので別バージョン

Version relations は、ある資源はもう一方の資源の過去の状態や編集で、同じ **Creator** によるものということ。

“「登録中の追悼時リリース版」 **is version of** “「エルトン・ジョンによる過去の同じタイトルの歌」”

- IsFormatOf / HasFormat

Title="Electronic AACR2" 電子版

Relation="Anglo-American Cataloging Rules, 2nd edition"紙媒体

[Relationship described is **IsFormatOf**] Format の違う同じもの。電子版が紙の別フォーマット (電子版(当該の資源)**is format of** 紙媒体 (関係資源))

Title="Landsat TM dataset of Arnhemland, NT, Australia" (ランドサット人工衛星が映した画像データセット)

Relation="arnhem.gif" (ランドサットのデータ画像の GIF 形式)

[Relationship described is **HasFormat**]

フォーマットの改変 Relation は、ある資源から別の資源が派生した場合。元の資源からの複製、再フォーマット技術。また、基本的に複製や再生産、再版などを目的としていて、新解釈ではないもの。

- IsReferencedBy/ References

Title="Morgan's Ancient Society" 『古代社会』ルイス・モルガン(1877)

Relation="Engels' Origin of the Family, Private Property and the State"『家族・私有 財産および国家の起源』エンゲルス(1884)

[Relationship described is **IsReferencedBy**]モルガンがエンゲルスに引用されている

Title="Nymphet Mania" ニンフマニア(1998)

Relation="References Adrian Lyne's 'Lolita'" エイドリアン・ライン監督「ロリータ」(1997)

[Relationship described is **References**] ニンフマニアがロリータを引用

Reference の関係は、ある資源の作成者が別の資源の引用、賛同、反論、主張をしいてるもの。

- Requires

Title="Dead Ringer" デッドリンガー (電子書籍)

Relation="Gemstar e-book" 読書用端末

[Relationship described is Requires] 電子書籍が, 端末機械を **Require**.

一方の資源が他方の資源を機能させるために必要とする。搬送媒体としてなど。関係資源が無ければ当該資源が使えない場合。

### 3.10.14. SHACHI: Utilization

Definition: 資源の用途情報

Comment: 当該資源は、研究・開発のフェーズにおいて、このように利用された。

※この項目は自動入力のため登録作業はしなくてもよい。

### 3.11. Term Name: **rights**

URI: <http://purl.org/dc/elements/1.1/rights>

Label: Rights 権利

Definition: 資源の著作権情報

Comment: 資源に関連する様々な財産権, 例えば知的財産権などの情報など。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※ 一行ずつ分けて, copy rights, access rights, license の 3 行を作る (情報があれば)

### 3.12. Term Name: **source**

URI: <http://purl.org/dc/elements/1.1/source>

Label: Source ソース (情報源)

Definition: 当該の資源の派生元資源

Comment: 当該の資源が関連資源からの全体あるいは部分的な派生である場合, 関連資源を公式の識別システムで指定することが望ましい。※SHACHI では資源名。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

※OLAC : 現在の資源はソース資源からの派生で全体部分関係にあるかもしれない。

※OLAC : 派生資源には既存資源を基にしているが, その場合知的財産権がある資源からの翻訳, ドラマ化, 編集, アノテーション付け, その他いろいろがある。例えばドラマがオリジナルであっても原作がソースであり, ドラマがオリジナル資源でありながら派生資源でもあるということになる。

著作権のあるものから作ったオリジナルが派生になるという考え方はダブリンコアでも **Source** として考えることになっているが, **Relation** 項目と重複する。なぜなら **isFormatOf** と **isVersionOf** のような **refinement** は技術的には派生資源の参照に使うから。

そこで、OLAC では Source 項目は、現在の資源の Type あるいは Creator あるいは Title がソース資源と随分違う場合に使うことにとどめる。オリジナル資源との違いが、Date, Format, Publisher, Identifier, Title の版の違いの場合は、Relation をエンコードにし、オリジナル資源との関係性を書く。

※ 例：

- ・ 小説や新聞から文例をとったコーパス＝小説や新聞の名前が Source
- ・ 文例をとった資源（元資源）から辞書を作った（派生資源）＝Relation

※ SHACHI の対応は OLAC に準じる。

Relation で特定できる関係資源は、優先的に Relation に入れ、Relation に入れたら Source に重複して入れないこと。知的内容の継承関係にある資源は特に Relation。知的内容を継承していない関係資源、例えば小説『赤毛のアン』から形容詞だけを抜き出した語彙集『赤毛のアン形容詞集』（架空）という言葉資源があった場合、原作の小説の知的内容を引き継いではいないので Source:『赤毛のアン』とする。一方、『赤毛のアン副詞集』が作られた場合、『赤毛のアン形容詞集』と『赤毛のアン副詞集』は Relation となる。

### 3.13. Term Name: **subject**

URI: <http://purl.org/dc/elements/1.1/subject>

Label: Subject 主題

Definition: 資源内容の主題

Comment: キーワード、キーフレーズ、分類コードなどを使って表示する。統制言語の使用推奨。場所や時間の主題を説明するには Coverage の方を使用。※SHACHI ではキーワード、キーフレーズを自由記入。

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

#### 3.13.1. OLAC: Linguistic Subject Vocabulary

特定の言語科学の下位分野で資源内容に関連するものを記述するのに使用

##### 3.13.1.1. anthropological\_linguistics

Definition: 言葉の形式や談話のプロセスから言語と文化の相関性を探り、言葉を通して文化を読み解こうと試みる言語学

Comment: 文献的には不完全なものが多い。民族言語学を含む。同一言語を話す複数の話し手のデータを収集しているものや、その言語が話されている地理的な地域

##### 3.13.1.2. applied\_linguistics

Definition: 言語学的方法論を使い言語の獲得・習得、言語教育を研究対象とする学問

Comment: 第2言語教授法や第2言語学習（TESOL や ESL）、対照言語学など。このカテゴリーはコンピュータのアプリケーションとは無関係、それは Computational Linguistics に分類する。

Example: 教材、教師用マニュアル、成人の語学学習の研究など

### 3.13.1.3. cognitive\_science:

**Definition:** 特に言語に関連した場合の人間の認知能力の研究

**Comment:**言語や視覚の実験, 心の研究に関する書籍

### 3.13.1.4. computational\_linguistics

**Definition:** コンピュータ科学を使用した言語学研究

**Comment:**人工知能や事前言語処理など。数理言語学ではない。

**Example:** 機械翻訳やテキスト, 発話のソフトウェア, 文章の構文分析などについての本や論文

### 3.13.1.5. discourse\_analysis

**Definition:** 自然発話における連続的な話し言葉の背後にあるパターンや意味の研究

**Comment:**ユーモア研究, 会話分析, 社会言語学のうち相互的なやりとりを扱うもの, 一定の環境における言語の調査, 例えば法廷における言語。

**Example:** 談話の文書化で, 話者の交代や重なりが記されているもの。談話表示理論の論文や様々な談話タイプ, 例えば医者と患者の会話, 研究に使える談話ビデオテープやオーディオ。

### 3.13.1.6. forensic\_linguistics

**Definition:** 法律分野における言語の応用

**Comment:** 法的な決定を下す際の言語方法論の使用を指す。法廷での言語は談話分析の方に分類するのが適している。

**Example:** 裁判訴訟における係争になる問題に関する論文。例えば原作者の特定, テキストのあいまいさの査定, 声の特徴など

### 3.13.1.7. general\_linguistics

**Definition:** 下位分野や特定の言語理論に特に拠らない特定のではない広範な言語学

**Comment:**下位分野を深くカバーする資源。ただし絶滅言語の詳細な統語的, 音韻的分析に関する論文のようなものは, **General linguistics** を使わず, それぞれの関連下位分野で特定していくべきである。

**Example:** 広範で導入的な教材, 例えばケンブリッジ言語百科事典や言語学用語の用語集など。

### 3.13.1.8. historical\_linguistics

**Definition:** 言語変化の通時的な研究

**Comment:**比較歴史言語学, 遺伝的分類, 文献学, 語源研究など。語源研究は **Lexicography** にも属する。

**Example:** ロマンズ語の母音の通時的な発達の研究。

### 3.13.1.9. history\_of\_linguistics

**Definition:** 言語科学の歴史と発達の研究

**Example:** ソシュールの伝記, プラトンの言語に関する論文。



**3.13.1.10. language\_acquisition**

**Definition:** 人間の言語の習得過程の研究

**Comment:** 子供と大人の両方の言語獲得に関連する教材資源, また母国語, 外国語両方の獲得に関する資源。しかし語学教授のために開発された資源, あるいは教育的な観点からの言語学習の過程を含むものは **Applied Linguistics** に分類。

**Example:** 母国語習得の研究, 言語習得実験のオーディオテープ, 言語獲得データ抽出の実験技術のマニュアルなど

**3.13.1.11. language\_documentation**

**Definition:** 言語説明と文書化の理論、方法、プロセス。

**Comment:** 言語説明と文書化の理論、方法、プロセス。フィールド方法論やフィールドワーク技術を含む。

**Example:** フィールドワーク技術のマニュアル

**3.13.1.12. lexicography**

**Definition:** 語彙資源や辞書のプロセスや収集や研究。単一言語、多言語とも。

**Example:** 辞書編纂についての本や論文。スウェーデン語彙リスト, アヴェスター語と英語のバイリンガル辞書, ある言語の法律用語集など。

**3.13.1.13. linguistics\_and\_literature**

**Definition:** フィクション、ドラマ、詩歌のような文学テキストへの言語学的分析の応用

**Comment:** 文体論と詩論を含む。

**Example:** 特定の著者の語彙パターン特徴の分析, あるいは文学的な効果を創造するために使われる統語的な構造。

**3.13.1.14. linguistic\_theories**

言語科学の基本的理論

**Comment:** 言語科学の基本的理論で、音韻論や統語論のような複数の下位分野に及ぶもの

**Example:** 普遍文法, 変形文法, モンタギュー文法を議論あるいは反論している論文

**3.13.1.15. mathematical\_linguistics**

言語の数学的な性質を扱う研究分野

**Comment:** 文法の形式的な特徴についての言語学論文

**3.13.1.16. morphology**

**Definition:** 語の内部構造について論じる学問分野

**Comment:** 形態的に複雑な言語からの一時的なデータを含む資源は, もし主要な目的が当該の言語の構造を記述することであれば, 'Language Description'に分類する。

Example: 形態論の理論についての論部, 『語文法』(Hudson) など。

### 3.13.1.17. neurolinguistics

Definition: 脳の構造と言語の脳内での言語の物理的な表示を研究する学問

Comment: 言語障害に関する脳の損傷についての研究。母国語習得の際の脳の神経の発達。後者は Language Acquisition としても分類すべきである。

### 3.13.1.18. philosophy\_of\_language

Definition: 言語と言語学理論への哲学の適用

Comment: サピア - ウォーフの仮説についての論文, 言語的思考と非言語的思考の比較。

### 3.13.1.19. phonetics

Definition: 発話の音の構造分類、個々の音、音の知覚などの研究

Comment: 調音音声学と音響音声学のどちらも、この定義に含まれる。データセット、語彙、実験結果のグラフによる表示などは、言語の音に特に関連性のあるものは Phonetics に分類する。

Example: 言語の音の一覧, 特定の音の調音の変数の研究。

### 3.13.1.20. phonology

Definition: ある言語のあるいは言語一般における音韻体系の背後にあるパターンと原理を研究する学問

Comment: 音韻分析に限らず, データセット, 語彙集, 実験結果のグラフ表示も, もし言語の特定の音韻体系に関連していれば Phonology に分類する。

Example: 音韻論は特定の言語に当てはまる。つまり言語の音節構造の研究, 音転呼の背景にある規則など。音韻論にはいるのは, 音素のデータセット, ある言語の音韻のフィールドノート, 音韻論を論じる論文たとえば最適性理論の議論など。

### 3.13.1.21. pragmatics

Definition: 言語使用を, それが話される文脈の点から研究する学問

Comment: 非言語的なコミュニケーションを含む。

Example: ポライトネス, 敬語表現, ダイクシス (直示), 発話行為など。

### 3.13.1.22. psycholinguistics

言語への心理学の適用。

Comment: 言語の背後にある心理的なメカニズム, 例えば記憶や知能の点から心理学を言語に応用した学問。(実験心理学)

Example: 特定の統語的制約の習得についての通言語的研究, 失語症の子どもの言語習得。

### 3.13.1.23. semantics

Definition: 言語構造の意味の研究

**Comment:** 会話における意味研究は談話分析に分類する。意味論では語のレベルを扱う（語彙意味論）

**Example:** 意味論の論文，例として，「心理条件意味論」。特定の言語における色の語彙集。

### 3.13.1.24. sociolinguistics

**Definition:** 有る社会で話される言語をその社会的背景において研究する言語学

**Comment:** 多言語使用，民族語，接触言語（ピジンやクレオール），言語地理学(方言学)，言語（標準化）計画，ジェンダー研究など

**Example:** 話し手の性別，年齢，社会層などの要因による言語変異の研究

### 3.13.1.25. syntax

**Definition:** 文の中で単語やその他の言語単位どうしの文法的な関係性を研究する分野。形態論は語の内部構造、統語論は文の内部構造であり混同しないこと

**Comment:** 特定の統語理論を使った言語の統語的説明。言語データを使って統語論を論じる論文。

### 3.13.1.26. text\_and\_corpus\_linguistics

**Definition:** 長い文章、テキスト、コーパスの言語学的な特徴の研究

**Comment:** テキストコーパスの計量分析や，記号論，ジャンル分析も。

**Example:** 英語話し言葉と書き言葉の法動詞の使用しついでの BNC の統計的分析。

### 3.13.1.27. translating\_and\_interpreting

**Definition:** ある言語を他の言語へ変換する行為の研究。口頭によるもの，文書によるもののどちらも含む。

**Comment:** 翻訳，通訳の質についてや，通訳支援の実用的なツールについての理論的な議論。翻訳されたテキストそのものの資源は，もしそれが言語構造を例示することを目的としていれば「Language Description」に分類される。

**Example:** レベルの違う翻訳の相対的なメリットについての研究。直訳，逐語訳など。

### 3.13.1.28. typology

**Definition:** 言語の系統に関係なく、通言語的な共通点相違点の研究をする学問。また結果的に言語を「言語の型」に分類する。

**Comment:** 普遍性の記述的で比較的な研究。

**Example:** 言語普遍性の文法の議論，あるいは反論としての複数言語のデータを使った論文。グリーンバーグの言語普遍性の点からの記述が進んでいない言語の分析。

### 3.13.1.29. writing\_systems

**Definition:** 話し言葉の紙やその他の媒体での視覚的な表示。書くこと、書記体系に含まれることを扱う。

**Comment:** 識字能力を扱った資源は書記体系か，応用言語学，あるいは両方に分類。その資源の大半が教育的な内容であるか（応用言語学），それとも書記素の体系の分析を含むのか（書記体系）による。

**Example:** 文字を持たない言語について、言語学者が書記体系を開発するなどの問題を考察する論文、絵文字の検証、暗号翻訳（クレタ線文字 B など）。

### 3.13.2. OLAC-Language extension

現存の言語あるいは絶滅言語として知られているすべての言語のコード  
ISO 国際標準化機構の言語コード

### 3.13.3. SHACHI: mono\_multi\_lingual

**Definition:** 単一言語資源か、多言語資源かを定める。

資源が **Subject**（対象）としている言語が、一種類であり、対応付けがないか、多種類あり、それぞれに対応付けがされていて、そのどちらも資源となるかを指定するために使用。

値は以下の 2 種類

#### 3.13.3.1. monolingual <SHACHI value>

**Definition:** 単一言語の資源。

**Comment:** 対象言語とその対象言語を説明する言語が同じ言語の場合。

**Example:** 毎日新聞 1991～2006 データファイル

#### 3.13.3.2. multilingual <SHACHI value>

**Definition:** 多言語の資源。

**Comment:** 対象言語が 2 つ以上の資源。または翻訳、多言語の対応を扱ったコーパス。

**Example:** ヨーロッパコーパスイニシアティブ多言語コーパス

### 3.13.4. SHACHI: Resource Subject

corpus, dictionary, thesaurus, glossary のどれかを選択

#### 3.13.4.1. corpus <SHACHI value>

**Definition:** 話し言葉、書き言葉などの収集物。

**Example:** 英語の話し言葉 100 万語コーパス、ルネサンス期の詩のコーパス、新聞データベースなど。

#### 3.13.4.2. dictionary <SHACHI value>

**Definition:** 辞書。ある言語の語のリストで同じ言語か別の言語でその語義や意味が付されているもの

**Example:** オクスフォード英語辞典（OED）、英辞郎 on the Web、学習英和辞典など

#### 3.13.4.3. thesaurus <SHACHI value>

**Definition:** 類義語辞書。ある言語の語のリストで語が似た意味のグループに分けられたもの。

**Example:** コウビルド類語辞典、オクスフォード学習類語辞典など

#### 3.13.4.4. glossary <SHACHI value>

Definition: コウビルド類語辞典, オクスフォード学習類語辞典など

Example: アメリカ合衆国政府のまとめた情報保証用語集, アスキー デジタル用語辞典, FOLDOP

### 3.14. Term Name: **title**

URI: <http://purl.org/dc/elements/1.1/title>

Label: Title タイトル

Definition: 資源名

Comment: 資源が公式にその名前によって知られている場合の, その名前

Type of Term: element

Status: recommended

Date Issued: 1999-07-02

#### 3.14.1. Term Name: alternative

URI: <http://purl.org/dc/terms/alternative>

Label: Alternative 代替名

Definition: 資源の公式タイトルに代わる他のあらゆる形式のタイトル

Comment: Title の省略形, 翻訳など。

Type of Term: element-refinement

Refines: <http://purl.org/dc/elements/1.1/title>

Status: recommended

Date Issued: 2000-07-11

※OLAC 例 タイトルと, その翻訳:

<title lang="x-sil-LLU">Na tala 'uria na idulaa diana</title>

<title refine="alternative" lang="en">The road to good reading</title>

※日本語資源の登録の際, タイトルと代替タイトルは以下の要領で入力

Title: Priority Areas "Spoken Dialogue" Simulated Spoken Dialogue Corpus

Title.alternative: PASD | 重点領域研究「音声対話」対話音声コーパス | Juten Ryo'iki Kenkyu "Onsei Taiwa" Taiwa Onsei Corpus

### 3.15. Term Name: **type**

URI: <http://purl.org/dc/elements/1.1/type>

Label: Type 資源タイプ

Definition: 資源の内容の特徴やジャンル

Comment: DMC1 の Type ボキャブラリなど, 統制言語を使用。ファイルのフォーマット (物理的な媒体や資源サイズ) を記述するときは Type ではなく Format を使用。

References: [DCMITYPE] <http://dublincore.org/documents/dcmi-type-vocabulary/>

Type of Term: element

Status: recommended

Date Issued: 1999-07-02 Type

※ 資源内容の一般的なカテゴリ, 機能, ジャンル, 集約レベルの説明用語を含む。Format 項目と混用しない。※SHACHI では四角で囲ったものを使用中。

### 3.15.1. The DCMI Type Vocabulary

記述例: <type code="Image"/>

- |   |   |
|---|---|
| * Collection  | * Service   |
| * Dataset   | * Software  |
| * Event   | * <span style="border: 1px solid black;">Sound</span> |
| * <span style="border: 1px solid black;">Image</span> | * <span style="border: 1px solid black;">Text</span>  |
| * Interactive Resource                                |   |

#### 3.15.1.1. Image

URI: <http://purl.org/dc/dcmitype/Image>

Label: Image イメージ

Definition: テキスト以外の視覚表示。

Comment: 物理的な対象の画像や写真。絵画, プリント, 描画, その他の画像やグラフィックス, アニメーション, 動画, 映画, 図表, 地図, ミュージカルの表記。電子のおよび物理的な表示の両方。

Type of Term: vocabulary-term

Broader Than: <http://purl.org/dc/dcmitype/StillImage>

Broader Than: <http://purl.org/dc/dcmitype/MovingImage>

Status: recommended

Date Issued: 2000-07-11

#### 3.15.1.2. Sound

URI: <http://purl.org/dc/dcmitype/Sound>

Label: Sound サウンド

Definition: 聞かれることを目的とした資源。

Comment: 音楽の再生ファイル形式, オーディオ CD, 録音された発話や音。

Type of Term: vocabulary-term

Status: recommended

Date Issued: 2000-07-11

#### 3.15.1.3. Text

URI: <http://purl.org/dc/dcmitype/Text>

Label: Text

Definition: 読まれるための文字を主に収録している資源

Comment: 本, 手紙, 詩, 新聞, 記事, 論文, メイリングリストのアーカイブなど。ファックス文書やテキストの画像もやはりこの Text に含まれる。

Type of Term: vocabulary-term

Status: recommended

Date Issued: 2000-07-11

### 3.15.2. Linguistic Data Types

[olac:linguistic-type]

- \* lexicon
- \* primary\_text
- \* language\_description

#### 3.15.2.1. lexicon

Name Lexicon

Definition 体系的な語彙項目のリストの資源

Comments 体系的な号項目のリストの資源を特定する。各語彙項目は定義、指示対象（固有名詞の場合など）、他の語彙項目との意味関係を示すものがあつたらいいが、なくてもいい。

Examples 例えば語彙リスト（対照言語表など）、類語辞典、WordNet、FrameNet、辞書（bilingual/multilingual）、用語辞典、固有名詞辞典などフレーズ集、イントネーション集など

#### 3.15.2.2. primary\_text

Name Primary Text

Definition 言語素材でそれ自体が研究対象となるもの。対象言語の未加工資源で、発話事象やそのような事象を書き留めたもの。

Comments たいてい一次テキストは時系列。実際の機能している言語事象。実際のあるいは架空の言語事象の時間的経過。会話、架空のシーン、発話誘導によるセッションなどで、アーカイブされているもの。

メモ：実際の媒体は問わない。媒体（録音機器など）が一次的な場合、ビデオ、カセットテープ、書面などであれば、その情報は **Format** に分類する。

注：**Primary\_text** は資源内容を指し物理的な形態ではない。もともと物理的媒体に記録されている。ビデオテープへの記録、カセットテープへの記録、書かれたテキストなど。物理的な形態は **Format** エレメントで指定。

Examples

時系列の文書化テキスト、手紙（録音されたもの）、誘導されたセッション（ビデオ化）、儀式やお話の読み聞かせセッション、その他録音された発話あるいは文書はなんでも、研究の対象となる言語構造を具現しているもの。

#### 3.15.2.3. language\_description

Name Language Description

Definition: 言語構造の体系的な文書化によって言語や言語の一面を記述した資源。

Comments

言語や、言語の特定の面を記述説明しているときに使う。他のものを説明している場合（エスペラントの創案者であるザメンホフを説明しているなど）は、当てはまらない。たいていは散文で、例文が随所にあるもの。

Examples: 形式文法, スケッチ文法, フィールドノート, 音韻スケッチ

### 3.15.3. Discourse Types

[olac:discourse-type]

およそ 10 種類の談話タイプを明記できる統制言語。Type の言語資源のジャンルを明確にするために使用(特に一次資源の場合)。Subject にも使える。資源が特定のジャンルであることを明記できる。

#### 3.15.3.1. drama

Definition: 台本があり、創作的で、演出された談話。2 人以上の participant がいる場合。

Examples: 演劇, スキット, 演じられたシーンなど

#### 3.15.3.2. formulaic\_discourse

Definition: 儀式的あるいは慣習的に構成された談話資源 v

Examples: 祈りの言葉, 祝福の言葉, 悪態, おまじない, 病気治癒の儀式, 結婚の誓いの言葉, 誓約など

#### 3.15.3.3. interactive\_discourse

Definition: 2 人以上の participant の間で交わされる言語を使った対話資源

Examples: 会話, インタビュー, 面接, 往復書簡などのやりとり, 相談, あいさつ, 別れのあいさつなど

#### 3.15.3.4. language\_play

Definition: もともと遊びの一部として機能する言語、あるいは発話の文体でその言語の構造に創造的な操作をしている資源

Examples: 表現法, ジョーク, 隠語, 偽装

#### 3.15.3.5. oratory

Definition: 公的な講話や慣習的な規則に従った演説方法での演説資源

Examples: 教会の説教, 講義, 政治的な演説, ミサなど

#### 3.15.3.6. narrative

Definition: 時間順に構成された出来事を表現した 1 人の語りの資源

Examples: 歴史的, 伝統的, 個人的な語り, 神話, 民話, 寓話, ユーモアなど

#### 3.15.3.7. procedural\_discourse

Definition: 方法、プロセス、順序を持つ状況などの説明や描写



Examples: レシピ, 解説, 計画など

### 3.15.3.8. report

Definition: ある出来事や状況の事実説明

Examples: ニュースレポート, エッセイ, 社説

### 3.15.3.9. singing

Definition: 音楽的な抑揚や声の抑揚を持つ言葉や音のひとまとまり(連続)。

Examples: チャンツ, 歌, コーラス

### 3.15.3.10. unintelligible\_speech

Definition: 発声から成る資源で、通常の言語として解釈されることを意図していない資源

Examples: 宗教的な言語, 統合失調症の患者の異言, 歌の一節 (ファララ ♪) など

## 3.15.4. SHACHI: Purpose

Definition: 言語資源を構築した目的

### 3.15.4.1. lexicography <SHACHI value>

Definition: 辞書編纂を目的に開発された資源

### 3.15.4.2. analysis <SHACHI value>

Definition: 分析を目的に開発された資源

### 3.15.4.3. developing\_technologies

Definition: 技術開発を目的に開発された資源

### 3.15.4.4. education

Definition: 教育での使用を目的に開発された資源

## 3.15.5. SHACHI: Style

Definition: 資源が話し言葉か書き言葉かどちらを収集しているか。

### 3.15.5.1. speech <SHACHI value>

Definition: 話し言葉の資源

Example: Longman Spoken American Corpus

### 3.15.5.2. written <SHACHI value>

Definition: 書き言葉の資源

Example: Longman Written American Corpus

### 3.15.6. SHACHI: Form

定型表現か非定型表現か、どちらを集めた資源であるかを選択(fixed, unfixed)

#### 3.15.6.1. fixed <SHACHI value>

Definition: 定型表現を集めた資源

Examples: Greetings: あいさつ: おはよう, さようなら。/ コンビニのレジの人の発話: いらっしゃいませ, 温めますか? / 手紙あいさつ文: 平素は格別のご高配を賜り、厚く御礼申し上げます。/ 式典のスピーチ定型句: 「本日は、お足元の悪い中お越しいただき、誠にありがとうございます。」 / 機械: 振込機や、Wordの手紙あいさつ文の自動入力など。法令文も含む。(書き方がある程度決められた性質のものであるため。)

#### 3.15.6.2. unfixed <SHACHI value>

Definition: 資源が集める表現形式に制限はない

Examples: 非定型文. 新聞データベース、同時通訳データベースなど(データに含まれる文が様々であるため)

### 3.15.7. SHACHI: Sentence

文の長さ

#### 3.15.7.1. short <SHACHI value>

Definition: 短い文のみ集めている

#### 3.15.7.2. long <SHACHI value>

Definition: 長い文のみ集めている

#### 3.15.7.3. mixed <SHACHI value>

Definition: 長い文と短い文が混ざっている

### 3.15.8. SHACHI: Annotation

Definition: 資源のテキストにつけられたタグ情報。ここでは特に資源を自動読み取り、解析する目的のタグ。

#### 3.15.8.1. annotated <SHACHI value>

Definition: 資源内容にアノテーションがついている

Example: ランカスター構文解析コーパス/フレーズマーカー(句ごとの標識)によって統語分析されている。/品詞ごと、形態素ごとにタグがついている、など

#### 3.15.8.2. plain <SHACHI value>

**Definition:** 資源内容にはアノテーションがついていない

### 3.15.9. SHACHI: Annotaion\_sample

**Definition:** アノテーションデータのサンプル

**Comment:** サンプルの貼り付け方

著作権に気をつける

原則全て貼り付ける

画像データの場合は URL

長い場合は 20 行だけ貼り付け，一行空けて”The rest is omitted”（以下省略）と記し，さらに一行空けて URL を貼る

### 3.15.10. SHACHI: Sample

**Definition:** 言語資源のサンプル

**Comment:** サンプルの貼り付け方は **Annotation** のサンプルの貼り方参照