

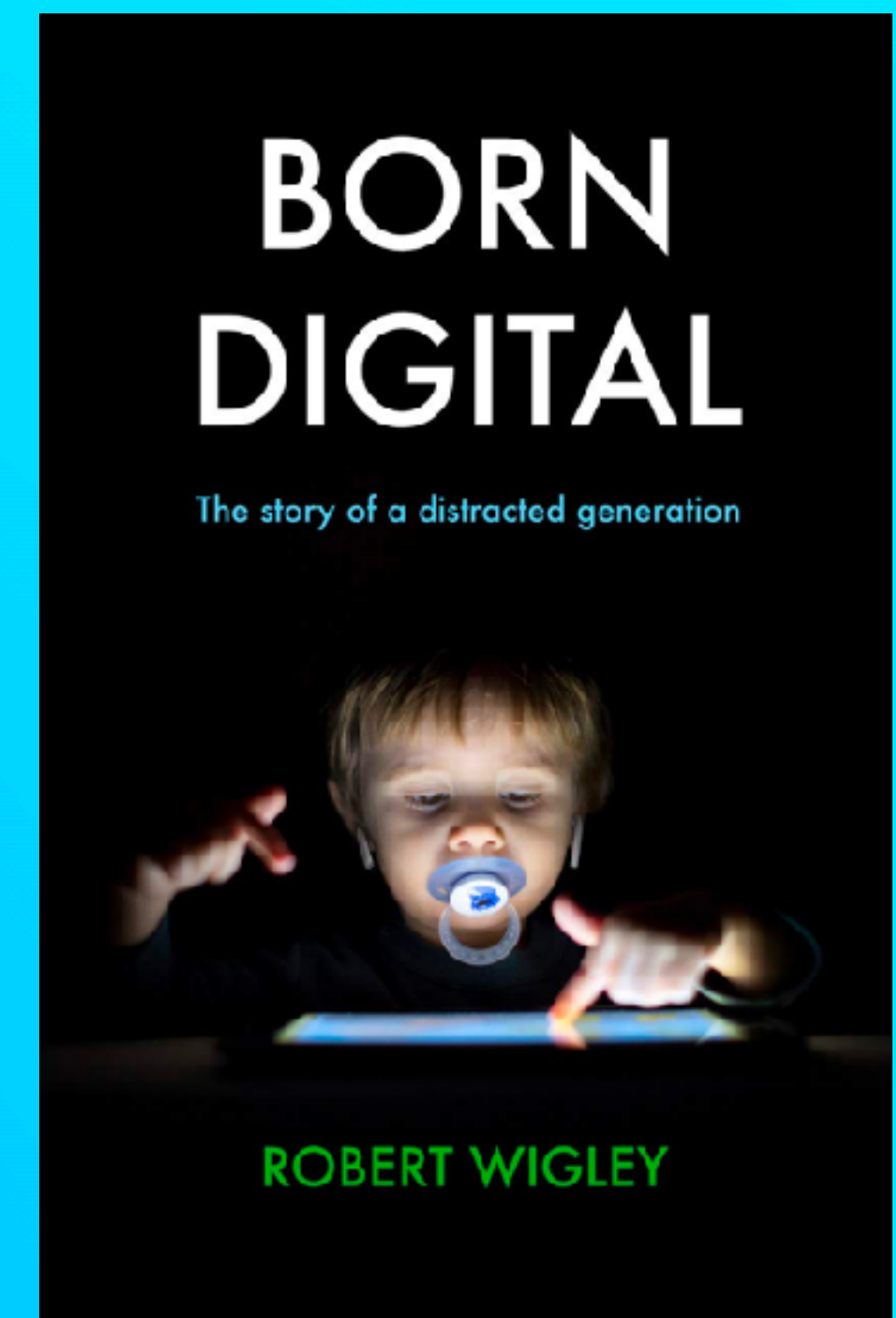
デジタルアーカイブの設計

日琉諸語のマイクロ類型論デジタルアーカイブ構築に向けて 加藤幹治（ROIS-DS人文
学オープンデータ共同利用センター／国立情報学研究所），宮川創（国立国語研究所）

宮川創 （国立国語研究所）

ボーンデジタル*な データセットのアーカイブ

*初めからデジタルでしか存在しないもののこと





World Atlas of Language Structures (WALS)

本プロジェクトのモデル

THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

Home Features Chapters Languages References Authors

Feature 7A: Glottalized Consonants

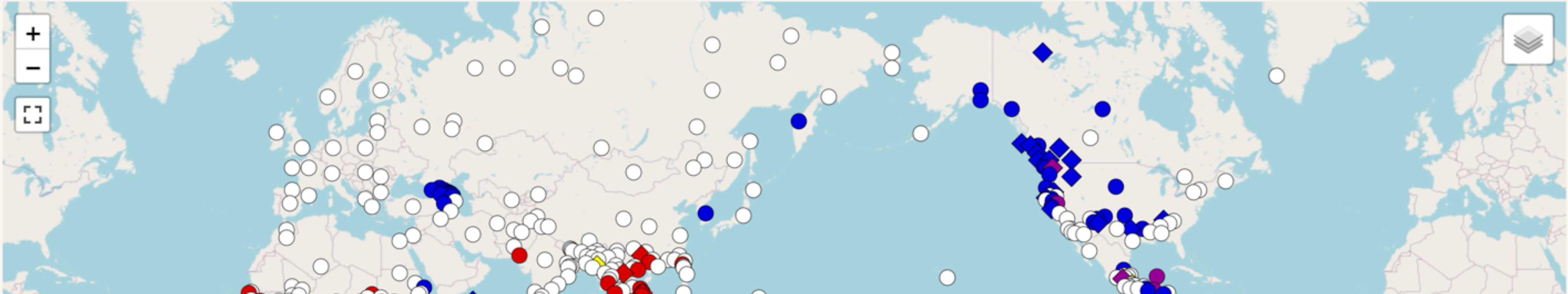
This feature is described in the text of chapter 7 [Glottalized Consonants](#) by [Ian Maddieson](#) [cite](#)

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

<input type="radio"/>	No glottalized consonants	409
<input checked="" type="radio"/>	Ejectives only	58
<input checked="" type="radio"/>	Implosives only	55
<input checked="" type="radio"/>	Glottalized resonants only	4
<input checked="" type="radio"/>	Ejectives and implosives	14
<input checked="" type="radio"/>	Ejectives and glottalized resonants	20
<input checked="" type="radio"/>	Implosives and glottalized resonants	4
<input checked="" type="radio"/>	Ejectives, implosives, and glottalized resonants	3

Legend Icon size Show/hide Labels GeoJSON



Cross-Linguistic Linked Data (CLLD) と CLDF (Cross-Linguistic Data Formats)

WALSを支えるテクノロジー

CLLD – Cross-Linguistic Linked Data

Home

News

About

Publication models

There are three major ways in which the CLLD project helps publishing cross-linguistic datasets:

1. Large databases may be published as standalone CLLD apps under the umbrella of the cldd edited series.
2. Smaller datasets may be submitted to one of the database journals started by the CLLD project.
3. Datasets may be hosted independently from the CLLD project, simply re-using the cldd software.

Published datasets

The following datasets are maintained by the CLLD project, i.e. fall into categories 1 and 2 above

Name	Description	Editors
WALS Online	The World Atlas of Language Structures	Matthew Dryer & Martin Haspelmath
WOLD	The World Loanword Database	Martin Haspelmath & Uri Tadon
APiCS Online	The Atlas of Pidgin and Creole Language Structures	Susanne Maria Michaelis, Phil Magnum Huber
ValPaL	Valency Patterns Leipzig	Iren Hartmann, Martin Haspelmath
eWAVE	The Electronic World Atlas of Varieties of English	Bernd Kortmann & Kerstin Lur
AfBo	A world-wide survey of affix borrowing	Frank Seifart
IDS	The Intercontinental Dictionary Series	Bernard Comrie & Hans-Jörg
ASJP	The database of the Automated Similarity Judgement Program	Søren Wichmann et al.
Numerals	Numerals in the World's Languages	Eugene Chan



Home

Specification

Ontology

Publications

Examples

The CLDF Ontology

Namespace: <http://cldf.cld.org/v1.0/terms.rdf#>

Version info: <http://cldf.cld.org/v1.2> (supersedes <http://cldf.cld.org/v1.1>)

Modules [↑](#)

Generic [↑](#)

<http://cldf.cld.org/v1.0/terms.rdf#Generic>

A generic CLDF dataset; i.e. a set of cross-linguistic data which does not fit any of the established CLDF modules.

Wordlist [↑](#)

<http://cldf.cld.org/v1.0/terms.rdf#Wordlist>

A dataset according to the [CLDF Wordlist](#) specification

Dictionary [↑](#)

<http://cldf.cld.org/v1.0/terms.rdf#Dictionary>

A dataset according to the [CLDF Dictionary](#) specification

StructureDataset [↑](#)

Modules

[Generic](#)

[Wordlist](#)

[Dictionary](#)

[StructureDataset](#)

[ParallelText](#)

Components

Properties

Reference Properties

`csvw:name`: "Generic"
`rdfs:subClassOf`: <http://www.w3.org/ns/dcat#Distribution>

`csvw:name`: "Wordlist"
`rdfs:subClassOf`: <http://www.w3.org/ns/dcat#Distribution>

`csvw:name`: "Dictionary"
`rdfs:subClassOf`: <http://www.w3.org/ns/dcat#Distribution>

`csvw:name`: "StructureDataset"

CLLD/CLDFの長所と短所

WALSを支える技術

長所

- 語彙, ライプチヒグロッシングルールでの例文の画一的記述, 表示, 地理情報の表示
- WALSとほぼ同じものが作れる

短所

- デザインや記述がやや単一的 (?)

Contents Mangement System (CMS)

他の方法1

- WordPress, Drupal, Joomla, PowerCMS, Omeka S など



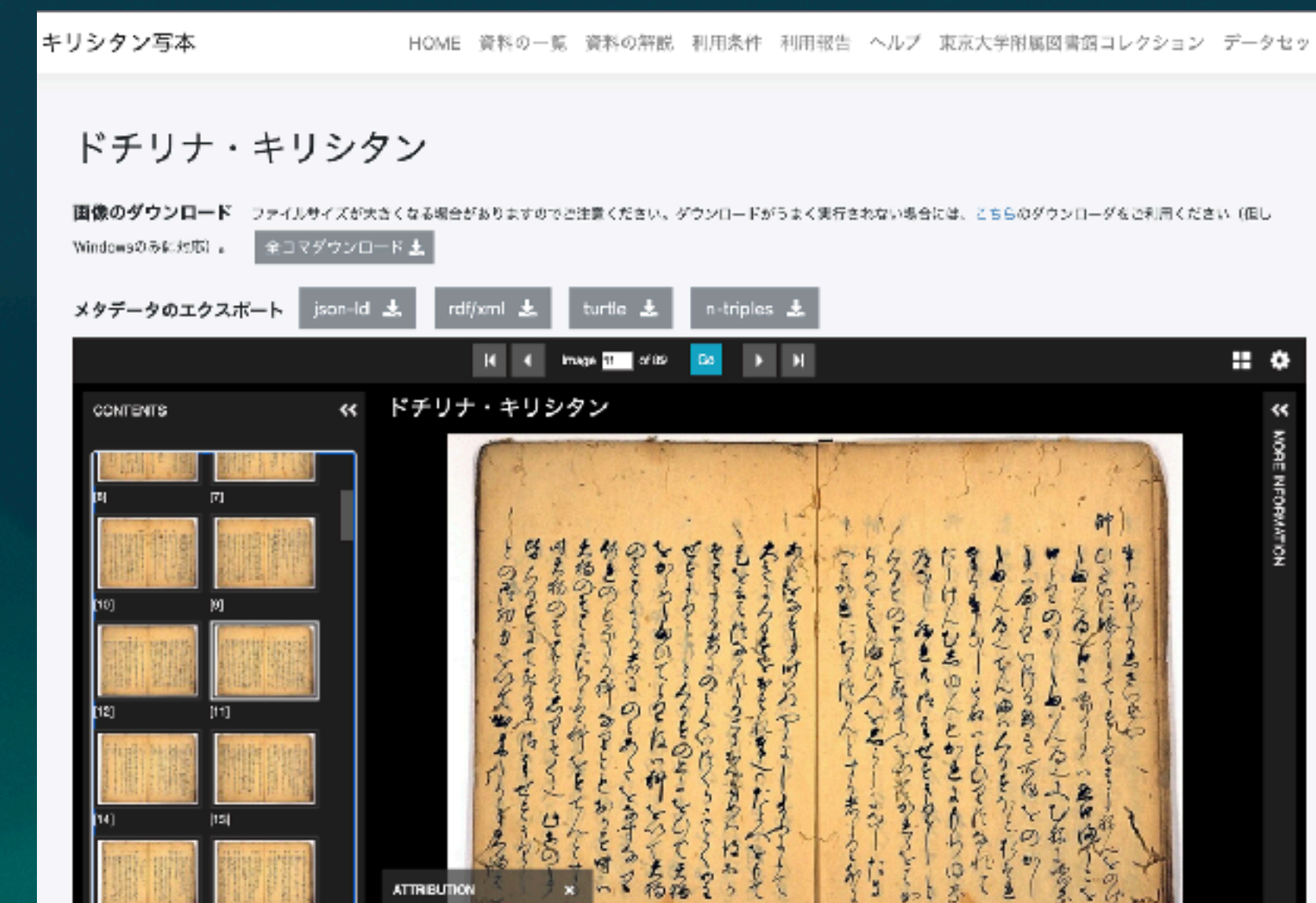
PowerCMSの例



WordPressの例



Drupalの例



Omeka Sの例

Static Site Generator (SSG)

他の方法2

- Gatsby, Hugo, Jekyllなど
- Markdownベースでの記事の記述
- レスポンスが速く、軽量のウェブサイト
- セキュリティが堅牢

The screenshot shows a dictionary entry for the Japanese word '叫びゆん' (Abiyun). The page includes a search bar at the top, a navigation menu, and a main content area with several sections. Annotations in blue boxes point to specific features:

- PoS** (Part of Speech): 自動詞 (Automatic verb)
- Conjugation type**: =raN, =ti
- Accent type**: ◎型
- Hiragana**: っあびゆん
- Kanji**: 叫びゆん
- Alphabet**: abiyun
- International Phonetic Alphabet (IPA)**: ?abijun
- Original**: ?abi=juN
- Sense 1**: 1. 叫ぶ。大声で呼ぶ。また、わめく。どなる。 (Call out loudly. Also, wameku. Do naru.)
- Sense 2**: 2. 大声で泣く。泣き叫ぶ。 (Cry loudly. Cry and call out.)
- Sense 3**: 3. ほえる。犬・猫・豚などが鳴く。 (Bark. Dog, cat, pig, etc. bark.)

The right sidebar shows the word 'っあび・ゆん' and its definitions:

- 1. 叫ぶ。大声で呼ぶ。また、わめく。どなる。
- 2. 大声で泣く。泣き叫ぶ。
- 3. ほえる。犬・猫・豚などが鳴く。

Hugoの例

現実的な方法？

国立国語研究所デジタルアーカイブNINDAのOmeka Sを再利用

The screenshot shows the Omeka S administration interface. On the left is a dark sidebar with navigation menus for 'admin', 'サイト一覧', 'リソース一覧', 'アイテム', 'アイテムセット', '語彙の一覧', 'リソーステンプレート一覧', '管理', 'ユーザー一覧', 'モジュール', 'ジョブ一覧', '設定', 'Assets', and 'モジュール' (with sub-items for CSV import, custom ontology, and item importer). The main content area is titled '新規サイト' (New Site) and contains a form with the following fields: 'タイトル' (Title) with the value 'LAJaR', 'URLスラッグ' (URL Slug) with the value 'lajar', and 'サムネイル' (Thumbnail) with a selection button. The footer of the interface displays 'Omeka S version 3.2.3 システム情報 ユーザーマニュアル レポートフォーラム'.

The screenshot shows a digital archive page from the National Institute of Japanese Language (NINDA). The page title is '方言録音資料シリーズ1「鹿児島市方言」1965年' (Dialect Recording Series 1 'Kagoshima City Dialect' 1965). The main content area features a large, faded speaker icon, indicating an audio player. Below the player, there is a metadata table:

タイトル	方言録音資料シリーズ1「鹿児島市方言」1965年
権利管理/Rights	CC0

The footer of the page includes logos for iif, H, UV, and a QR code.