

DCAT と CKAN

武田英明

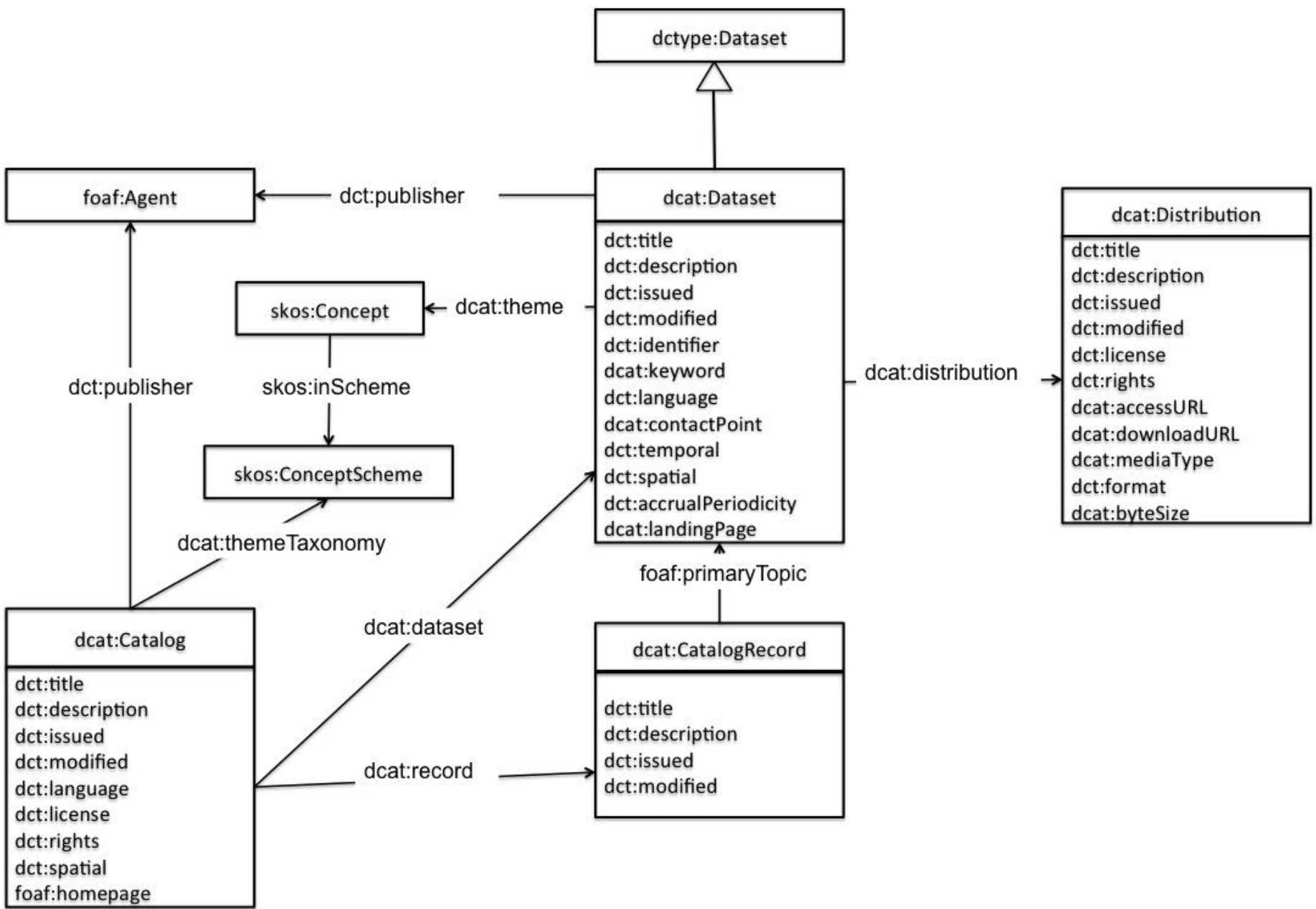
国立情報学研究所

takeda@nii.ac.jp

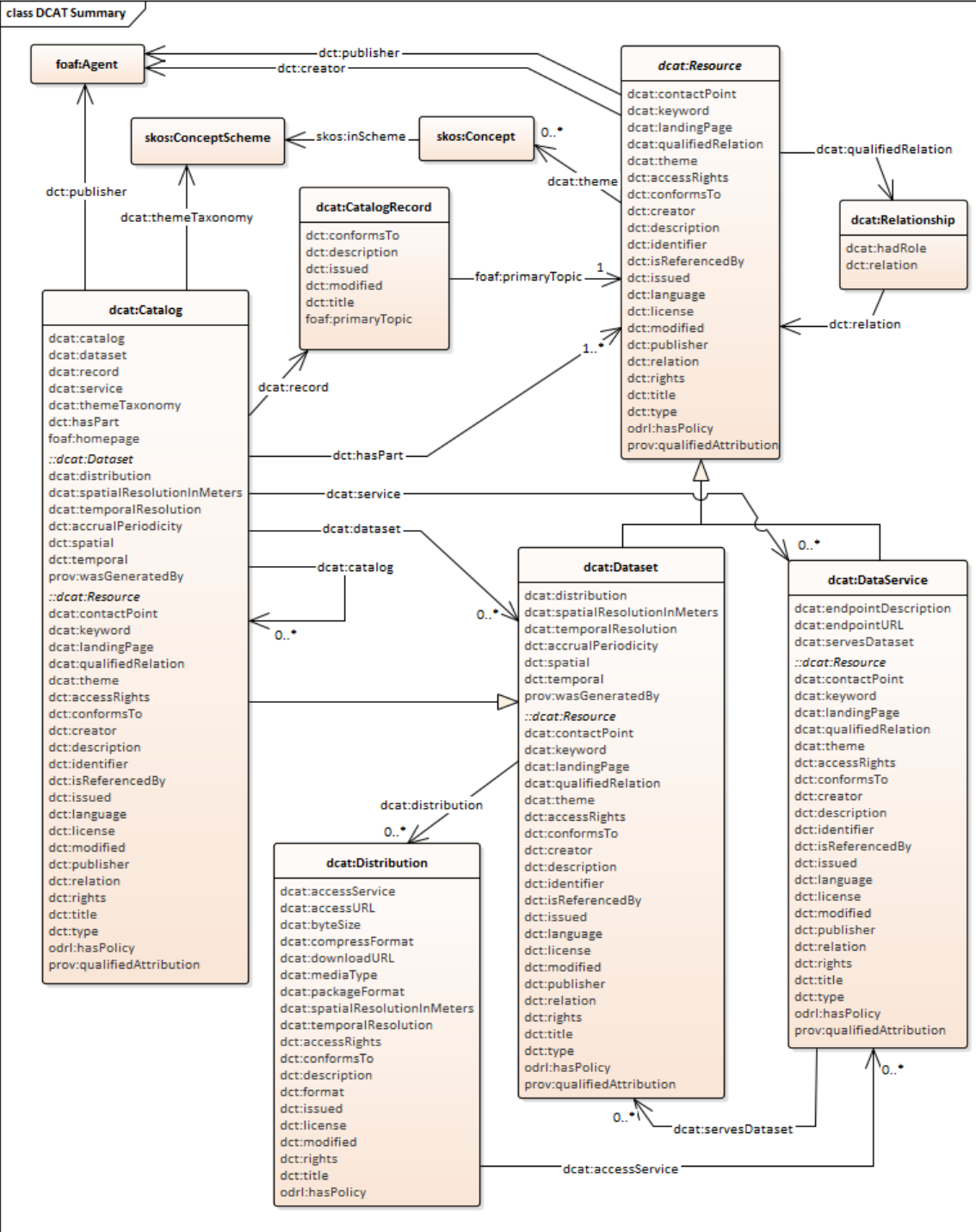
DCAT(Data Catalog Vocabulary)

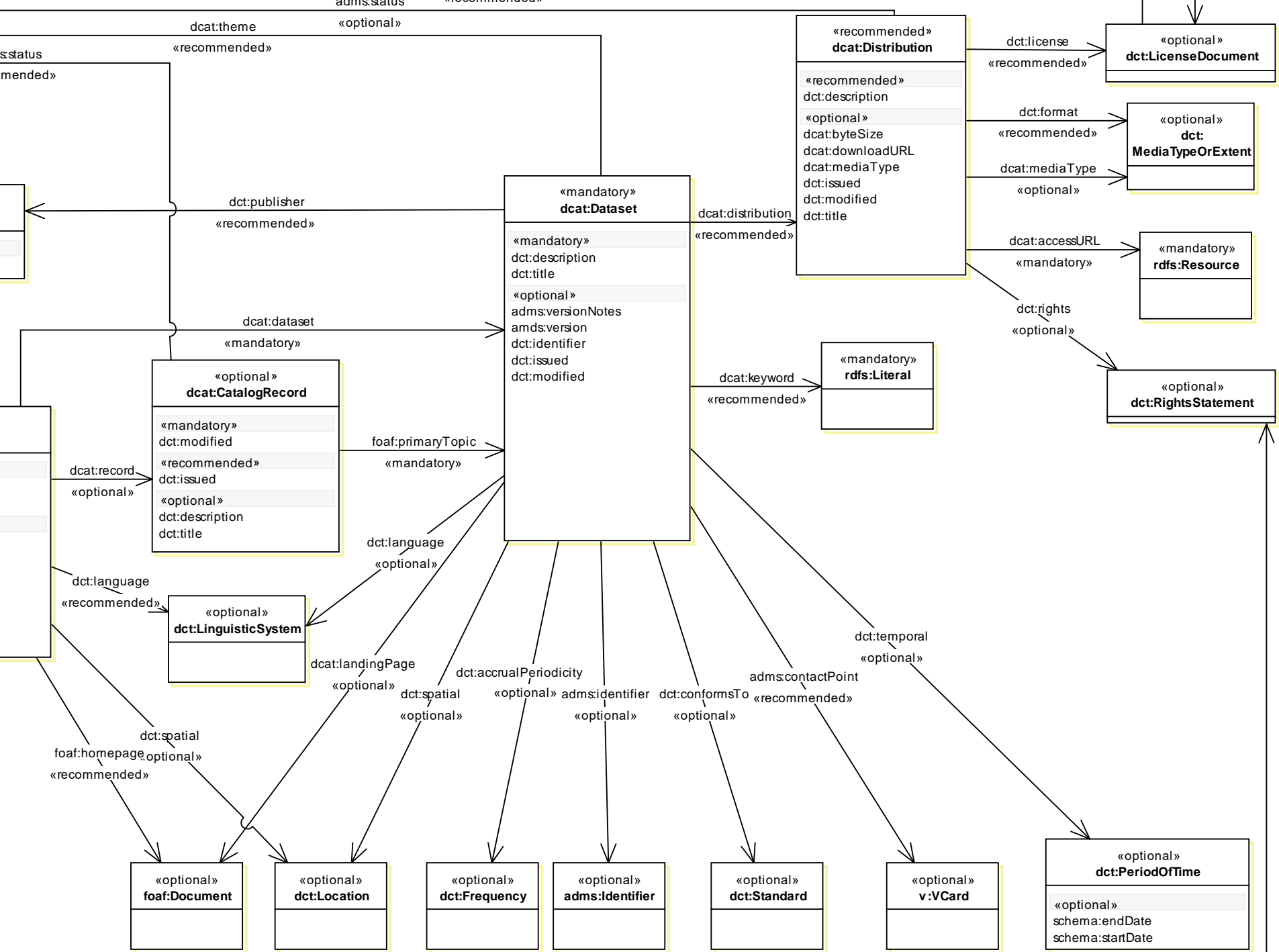
- Web上のデータカタログの相互運用性のための RDF語彙
- 経緯
 - Semantic Webのコミュニティから提案
 - W3Cの規格に
 - EUのデータ公開での規格へ (DCAT-AP)
- W3C勧告
 - 第1版 2014年1月16日
 - 第2版 2020年2月4日

DCAT 基本構造 (version 1)



DCAT 基本構造 (version 2)





Comprehensive Knowledge Archive Network (CKAN)

- オープンデータの保管と公開のためのオープンソースのデータポータルソフトウェア
 - Open Knowledge Foundationが管理
- 世界の主要なオープンポータルで採用
 - data.gov, data.go.uk,
- 基本構成：PostgreSQL(DB), Apach Solr(検索)
- Web UI, API(入力、出力)

CKANへようこそ

これはCKANあるいはサイト全般についての良い紹介文です。我々はまだここへ行くためのコピーがありませんが、すぐに行くでしょう。

420 x 220

これは特徴のセクションです

データ検索

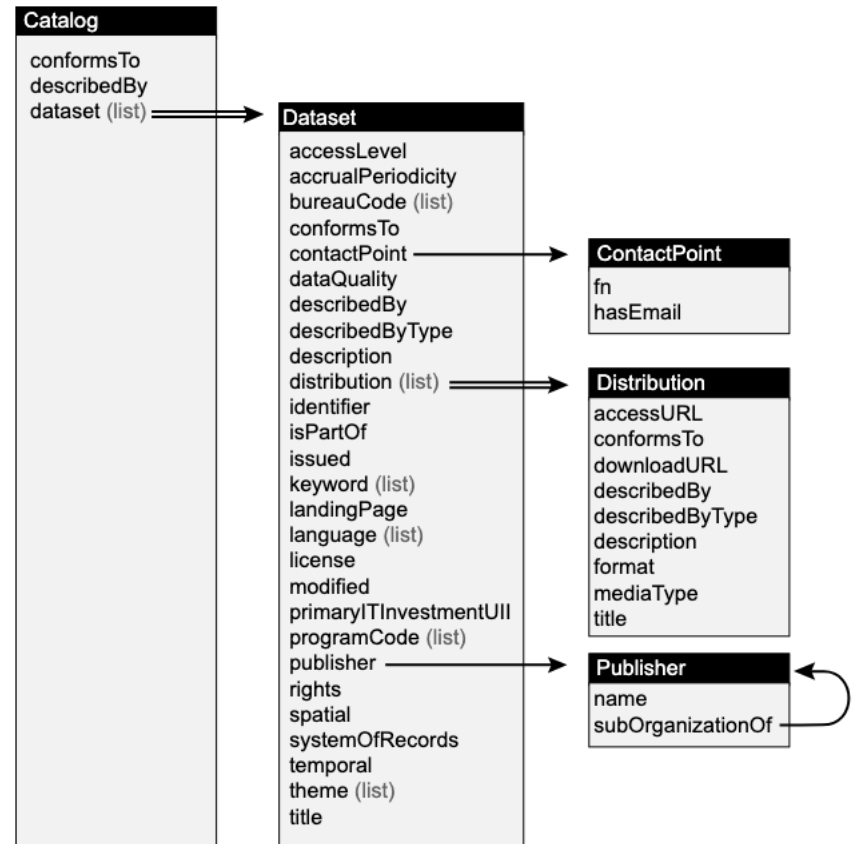
例: 環境



人気のあるタグ

CKAN と DCAT

- 基本構造
 - Catalog– Dataset – Distribution
- 各クラスのデータ項目は任意に設定可能
- extras



data.govの例

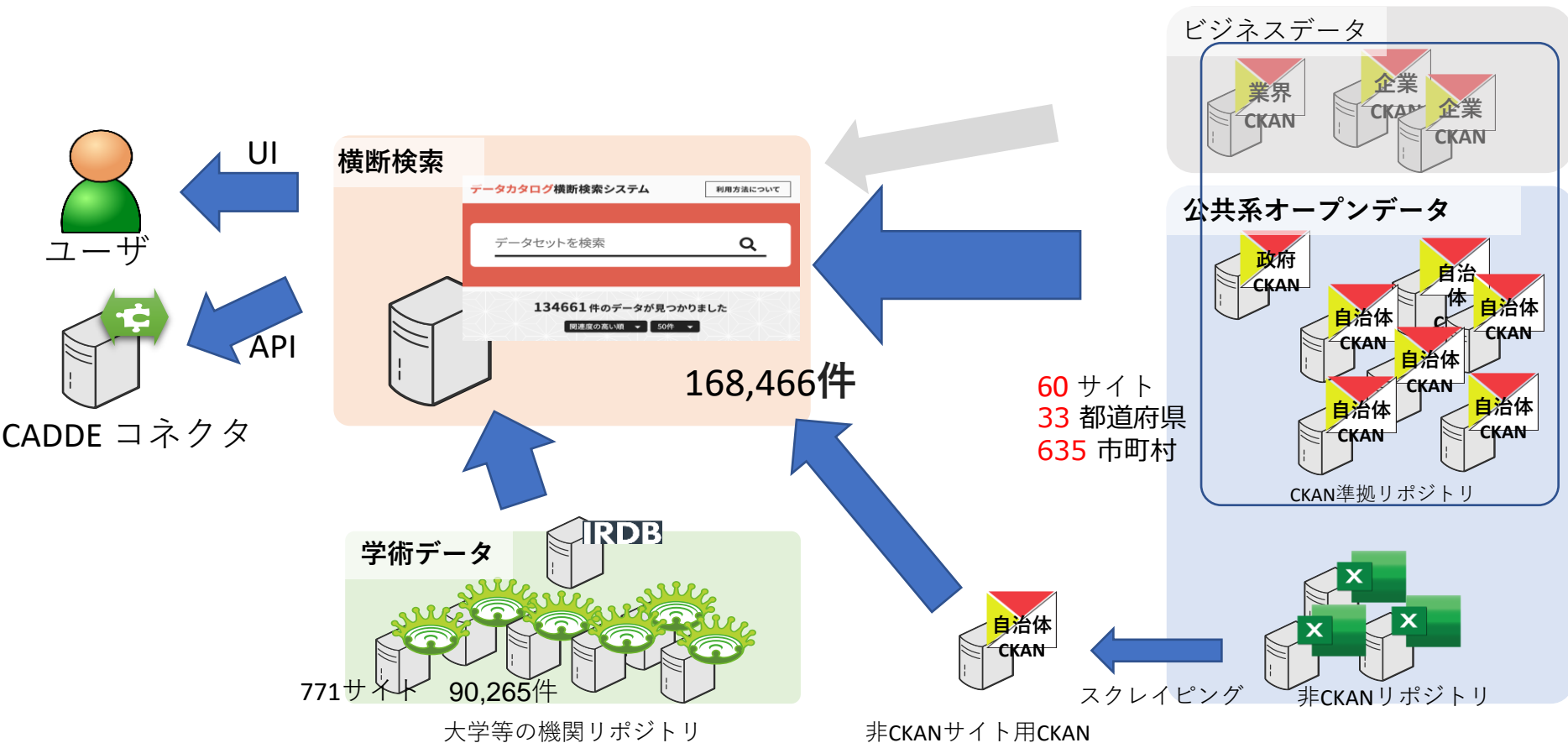
DCAT-CKAN

- CKANデータ項目をDCATとして出力
 - DCATのExtension (ckanext-dcat)

DCAT class	DCAT property	CKAN dataset field
dcat:Dataset	-	extra:uri
dcat:Dataset	dct:title	title
dcat:Dataset	dct:description	notes
dcat:Dataset	dcat:keyword	tags
dcat:Dataset	dcat:theme	extra:theme
dcat:Dataset	dct:identifier	extra:identifier
dcat:Dataset	adms:identifier	extra:alternate_identifier
dcat:Dataset	dct:issued	extra:issued
dcat:Dataset	dct:modified	extra:modified
dcat:Dataset	owl:versionInfo	version
dcat:Dataset	adms:versionNotes	extra:version_notes
dcat:Dataset	dct:language	extra:language
dcat:Dataset	dcat:landingPage	url
dcat:Dataset	dct:accrualPeriodicity	extra:frequency
dcat:Dataset	dct:conformsTo	extra:conforms_to
dcat:Dataset	dct:accessRights	extra:access_rights
dcat:Dataset	foaf:page	extra:documentation
dcat:Dataset	dct:provenance	extra:provenance
dcat:Dataset	dct:type	extra:dcat_type
dcat:Dataset	dct:hasVersion	extra:has_version
dcat:Dataset	dct:isVersionOf	extra:is_version_of
dcat:Dataset	dct:source	extra:source
dcat:Dataset	adms:sample	extra:sample
dcat:Dataset	dct:spatial	extra:spatial_uri
dcat:Dataset	dct:temporal	extra:temporal_start + extra:temporal_end
dcat:Dataset	dct:publisher	extra:publisher_uri
foaf:Agent	foaf:name	extra:publisher_name
foaf:Agent	foaf:mbox	extra:publisher_email
foaf:Agent	foaf:homepage	extra:publisher_url
foaf:Agent	dct:type	extra:publisher_type
dcat:Dataset	dcat:contactPoint	extra:contact_uri
vcard:Kind	vcard:fn	extra:contact_name
vcard:Kind	vcard:hasEmail	extra:contact_email
dcat:Dataset	dcat:distribution	resources
dcat:Distribution	-	resource:uri
dcat:Distribution	dct:title	resource:name
dcat:Distribution	dcat:accessURL	resource:access_url
dcat:Distribution	dcat:downloadURL	resource:download_url
dcat:Distribution	dct:description	resource:description
dcat:Distribution	dcat:mediaType	resource:mimetype
dcat:Distribution	dct:format	resource:format
dcat:Distribution	dct:license	resource:license
dcat:Distribution	adms:status	resource:status
dcat:Distribution	dcat:byteSize	resource:size
dcat:Distribution	dct:issued	resource:issued
dcat:Distribution	dct:modified	resource:modified
dcat:Distribution	dct:rights	resource:rights
dcat:Distribution	foaf:page	resource:documentation
dcat:Distribution	dct:language	resource:language
dcat:Distribution	dct:conformsTo	resource:conforms_to
spdx:Checksum	spdx:checksumValue	resource:hash
spdx:Checksum	spdx:algorithm	resource:hash_algorithm

データカタログ横断検索

- ✓ 組織（企業、団体、自治体、大学など）が公開するデータカタログサイトからカタログ情報を収集して、検索可能にする。
- ✓ ユーザによるWeb UIを通じた簡易検索/詳細検索、CADDEコネクタによるAPI利用



数字は2023年3月1日現在

<https://search.ckan.jp>

COVID



71件のデータが見つかりました

関連度の高い順

50件

 個別表示 シリーズをグループ化 重複結果をグループ化

検索条件

条件クリア ×

組織

北九州市 (8)

福岡県 (8)

山梨県 (7)

札幌市 (7)

福岡市 (5)

> 全て表示

サイト名

BODIK ODCS (27)

DATA-SMART CITY SAPPORO (8)

ふじのくにオープンデータカタログ (8)

山梨県オープンデータサイト (7)

学術機関リポジトリ (4)

> 全て表示

タグ (複数選択可)

健康 (15)

医療 (14)

新型コロナウイルス関連

BODIK ODCS



大分県内の新型コロナウイルス感染症に関する情報です。/ 更新は不定期です。最新の情報は大分県公式サイト以下のページを参照してください。/ <https://www.pref.oita.jp/site/covid19-oita/covid19-pcr.html>【リソース】新型コロナ陽性事件数 / 登録日は最終更新日参照のこと。更新は不定期です。全数把握見直しに伴い、令和4年9月27日以降は一部のみの公開となります。検査数は令和4年11月30日までとなります。最新の情報は大分県公式サイト以下のページを参照してください。/ <https://www.pref.oita.jp/site/covid19-oita/covid19-pcr.html> / 患者属性情報 / 令和2年3月3日～令和4年9月26日までの患者属性情報を公開しています。/ 全数届出の見直し以降(令和4年9月27日以降)の更新はありません。/ 本県における最新の報告については以下のページをご確認ください。/ <https://www.pref.oita.jp/site/covid19-oita/covid19-pcr.html>【キーワード】健...

最終更新日: 2023/02/22 16:01:28 | 📄: CSV

新型コロナウイルス感染症県内感染動向

ふじのくにオープンデータカタログ

県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数【リソース】220001_shizuoka_covid19_test_number.csv / 220001_new_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_call_center.csv

最終更新日: 2023/02/07 01:29:24 | 📄: CSV

新型コロナウイルス感染症 (COVID-19) が心理臨床業務に与えた影響についての調査データ

学術機関リポジトリ

このデータは、新型コロナウイルス感染症 (COVID-19) が心理臨床業務に与えた影響について実施されたwebアンケートの結果である。実施時期は、緊急事態宣言が出された直後の2020年5月17日から6月16日までで、有効回答者は465名である。/ This data is the result of a web survey on the impact of COVID-19 on psychological practice. The period of the survey was from May 17 to June 16, 2020, just after the declaration of the state of emergency, and there were 465 valid participants. / Keywords: COVID-19, Clinical Psychology, Psychological Practice, Japan【リソース】Fulltext

最終更新日: 2022/03/23 00:27:52 | 📄: txt/csv

新型コロナウイルスの検査状況等について

愛媛県オープンデータ

<https://search.ckan.jp>

データセットを検索



160135件のデータが見つかりました

更新日

50件

 個別表示 シリーズをグループ化 重複結果をグループ化

検索条件

条件クリア ×

組織

IRDB (77302)

山梨県 (10477)

経済産業省 (2949)

厚生労働省 (2587)

国土交通省 (2508)

> 全て表示

サイト名

学術機関リポジトリ (77302)

DATA GO JP データカタログ
サイト (22220)

G空間情報センター (12054)

山梨県オープンデータサイ
ト (10477)

BODIK ODCS (10353)

> 全て表示

タグ (複数選択可)

統計 (10478)

予算_決算_調達関連情報 (4581)

統計調査結果 (4561)

財政 (2642)

白書_年次報告 (2445)

> 全て表示

理容所台帳

ふじのくに オープンデータカタログ

台帳登録されている理容所の一覧です。【リソース】R2.2理容所新規許可施設 / R2.3理容所新規許可施設 / R2.1理容所新規許可施設 / 理容所許可全施設 (R2.3.31現在) / R2.5理容所新規許可施設 / R2.7理容所新規許可施設 / R2.11理容所新規許可施設 / 理容所許可全施設 (R3.3.31現在) / R3.4理容所新規許可施設 / R3.6理容所新規許可施設

最終更新日: 2023/06/22 09:18:56 | 📄: CSV

PRTRデータ

ふじのくに オープンデータカタログ

PRTR届出に関する集計結果【リソース】PRTR概要2013 / PRTR概要2016 / PRTR物質別2016 / PRTR物質別2013 / PRTR概要2014 / PRTR物質別2014 / PRTR業種別2014 / PRTR概要2015 / PRTR物質別2015 / PRTR業種別2013 / PRTR業種別2015 / PRTR業種別2016 / PRTR業種別2017 / PRTR業種別2018 / PRTR業者別2019 / PRTR業者別2020 / PRTR業者別2021

最終更新日: 2023/06/22 09:06:57 | 📄: DOCX, XLSX, XLS

東京都防災マップ 避難所・避難場所一覧データ

東京都オープンデータカタログサイト

【東京都総務局】東京都防災マップ(<https://map.bosai.metro.tokyo.lg.jp/index.html>)に掲載している避難所、避難場所の一覧です。なお、本情報は、原則として令和4年4月1日基準での情報です。 / データ利用に当たっては、東京都提供の「避難所、避難場所データ オープンデータ」を利用していることを表示してください。 / 本データに関するお問い合わせについては、東京都総務局総合防災部防災計画課(電話:03-5388-2454)までお問い合わせください。 / <ライセンス> / 本データはクリエイティブ・コモンズ・ライセンス表示 4.0 国際 (<https://creativecommons.org/licenses/by/4.0/legalcode.ja>) の下に提供いたします。また、利用者がOpen Data CommonsによるODC BY (<https://opendatacommons.org/licenses/by/1-0/>) 又はODbL (<https://opendatacommons.org/licenses/odbl/>) の利用を希望する場合に、それを妨げるもので...

最終更新日: 2023/06/22 09:06:49 | 📄: CSV, XLSX

静岡市オープンデータサイト利用状況(～令和3年度分)

ふじのくに オープンデータカタログ

静岡市オープンデータサイトやデータファイルの月別のアクセス数の状況です。(令和3年度分まで) / H27.12月利用状況から下記の条件でのデータに変更しました。 / ・ロボットによるアクセスは除外しております。 / ・検索サイト等からのデータへの直リンクによるアクセスは含まれておりません。 / ・アクセス集計とリソースダウンロード数の集計処理の違いにより数値にズレが発生している場合があります。【リソース】オープンデータサイト利用状況_csv / オープンデータポータルサイト、カタログサイト及びデータファイルのアクセス数 ※2015.12からデータカウント方法を変更。・ロボットによるアクセスは除外しました。・検索サイト等からのリソースへの直リンクによるアクセスは含まれておりません。・アクセス集計とリソースダウンロード数の集計処理の違いにより数値にズレが発生している場合があります。 / オープンデータサイト利用状況_json / オープンデータサイト利用状況_rdf / H27.11月期オープンデータアクセスTOP10_csv / H27.11月期にアクセスされた上位10データ。 / H27.11月期オープ...

最終更新日: 2023/06/22 08:50:49 | 📄: CSV, JSON, RDF

新型コロナウイルス感染症県内感染動向

<https://ckan.pref.shizuoka.jp/dataset/852e2fee-671c-4067-a914-bb77d3a>

「ふじのくに オープンデータカタログ」に移動

データセットについて

県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数【リソース】220001_shizuoka_covid19_test_number.csv / 静岡県内における新型コロナウイルス感染症の検査件数 / 220001_new_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_call_center.csv



データセット

xckan_id	ckan.pref.shizuoka.jp__dataset:852e2fee-671c-4067-a914-bb77d3a
xckan_title	新型コロナウイルス感染症県内感染動向
xckan_site_name	ふじのくに オープンデータカタログ
xckan_site_url	https://ckan.pref.shizuoka.jp/dataset/852e2fee-671c-4067-a914-bb77d3a
xckan_last_updated	2023-03-11T11:24:05Z
xckan_description	県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数【リソース】220001_shizuoka_covid19_test_number.csv / 静岡県内における新型コロナウイルス感染症の検査件数 / 220001_new_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_call_center.csv
xckan_tags	静岡県
データセットのライセンス	クリエイティブ・コモンズ 表示
maintainer	
relationships_as_object	
private	false
maintainer_email	
num_tags	0
id	77c5ea94-9956-4af0-b41f-314b50e7949d
データセットの発行日(自動設定)	2023-03-10T11:29:53.041430
データセットの更新日または修正日(自動設定)	2023-03-11T11:24:05.675391

author	健康福祉部感染症対策局新型コロナウイルス対策課
author_email	
state	active
version	
creator_user_id	035cb0d7-731e-44af-beaa-dff7e8d2023b
type	dataset
num_resources	4
license_id	cc-by
relationships_as_subject	
name	852e2fee-671c-4067-a914-bb77d3a
isopen	true
データセットの説明ページURL	
データセットの説明	県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数【リソース】220001_shizuoka_covid19_test_number.csv / 静岡県内における新型コロナウイルス感染症の検査件数 / 220001_new_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_call_center.csv
owner_org	fe8ce972-748f-4781-b010-321000000000
データセットのライセンス(URL)	http://www.openstreetmap.org/licenses/by
データセットのタイトル	新型コロナウイルス感染症県内感染動向
revision_id	95aae62e-44d3-4067-a914-bb77d3a
score	6.559884

組織

description	
created	2018-12-05T09:17:00Z
title	静岡県
name	org1
is_organization	true
state	active
image_url	
revision_id	5d063b46-8451-4067-a914-bb77d3a
type	organization
id	fe8ce972-748f-4781-b010-321000000000
approval_status	approved

グループ

- グループ1

display_name	e-Stat分類-保険衛生
description	
image_display_url	
title	e-Stat分類-保険衛生
id	62ec8724-5e35-4fb4-ba17-04bed22a790d
name	ssg160

- グループ2

display_name	静岡県-健康づくり・疾病対策と感染症の予防
description	
image_display_url	
title	静岡県-健康づくり・疾病対策と感染症の予防
id	fbada0c7-6cc8-4ca8-834d-c4e77b75439c
name	ssg54

配信

- 配信1

cache_last_updated	
package_id	77c5ea94-9956-4af0-b41f-314b50e7949d
webstore_last_updated	
datastore_active	false
id	bb83fb96-bdbd-429d-bd55-9b72e5edaa4a
配信のバイトサイズ	
state	active
hash	

配信の説明	静岡県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数【リソース】220001_shizuoka_covid19_test_number.csv / 静岡県内における新型コロナウイルス感染症の検査件数 / 220001_new_shizuoka_covid19_patients.csv / 220001_shizuoka_covid19_call_center.csv
配信の最終変更日(自動設定)	2023-03-10T11:29:54.716770

<https://search.ckan.jp>

グループ

新型コロナウイルス感染症県内感染動向

組織



静岡県

この組織には説明がありません

ソーシャル

Google+

Twitter

Facebook

ライセンス

クリエイティブ・コモンズ 表示

データセット

グループ

アクティビティストリーム

新型コロナウイルス感染症県内感染動向

県内における新型コロナウイルス感染症の感染動向、検査人数、相談件数

データとリソース



220001_shizuoka_covid19_test_number.csv

静岡県内における新型コロナウイルス感染症の検査件数

[詳細](#)



220001_new_shizuoka_covid19_patients.csv

[詳細](#)



220001_shizuoka_covid19_patients.csv

[詳細](#)



220001_shizuoka_covid19_call_center.csv

[詳細](#)

追加情報

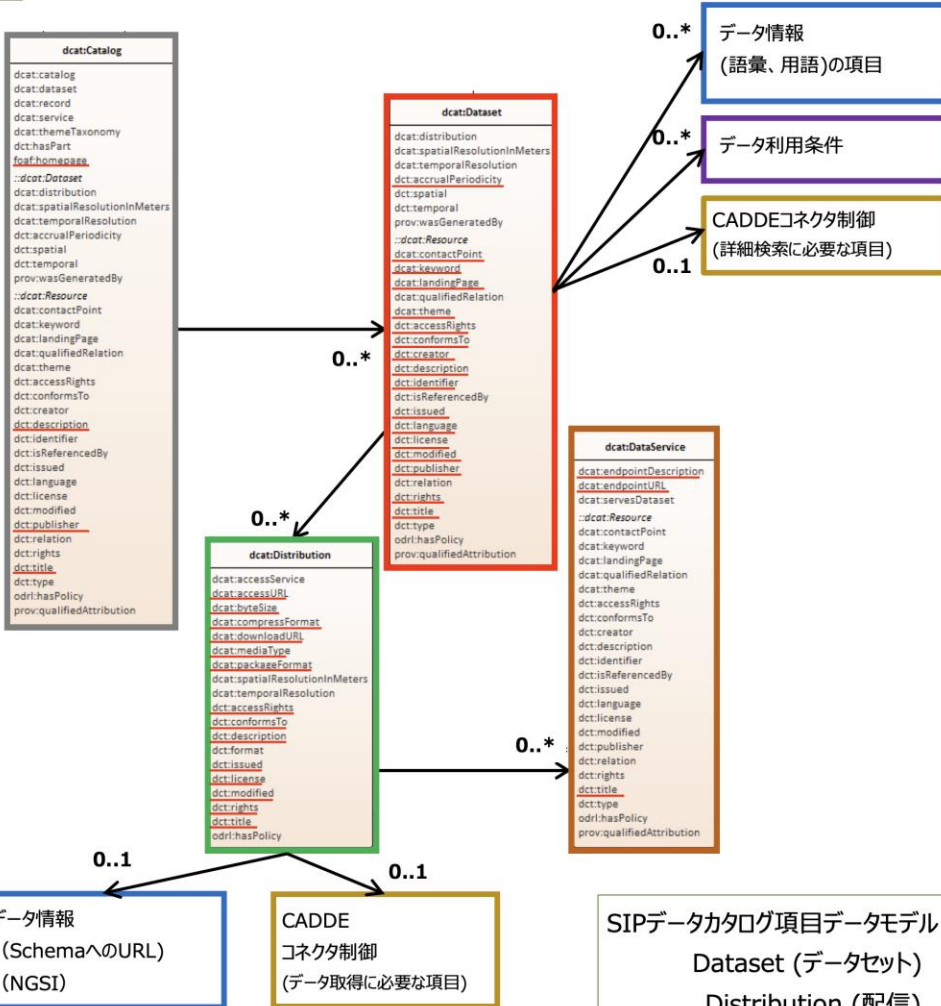
フィールド	値
組織	静岡県
作成者	健康福祉部感染症対策局新型コロナウイルス対策課
作成日	2023-03-10 20:29

SIPプロジェクトでのDCAT拡張

SIPとして選定したW3C DCAT 2.0のデータモデルに対する拡張

SIPデータ
カタログ
項目仕様

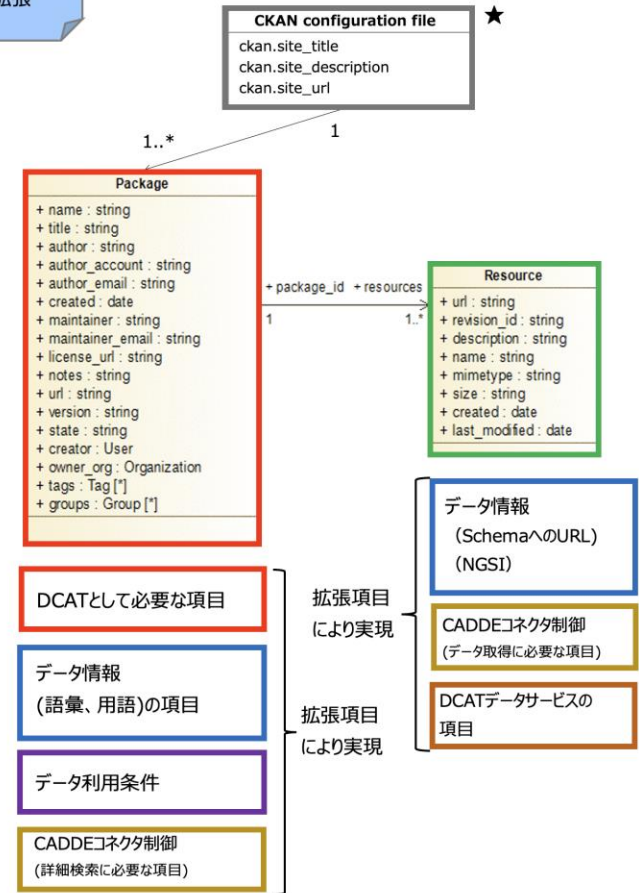
SIP データカタログ項目 Data Model
(DSAベース: W3C DCAT, 利用条件他)



CKANのデータモデル

CKAN
データ項目
拡張

CKAN Data Modelの拡張



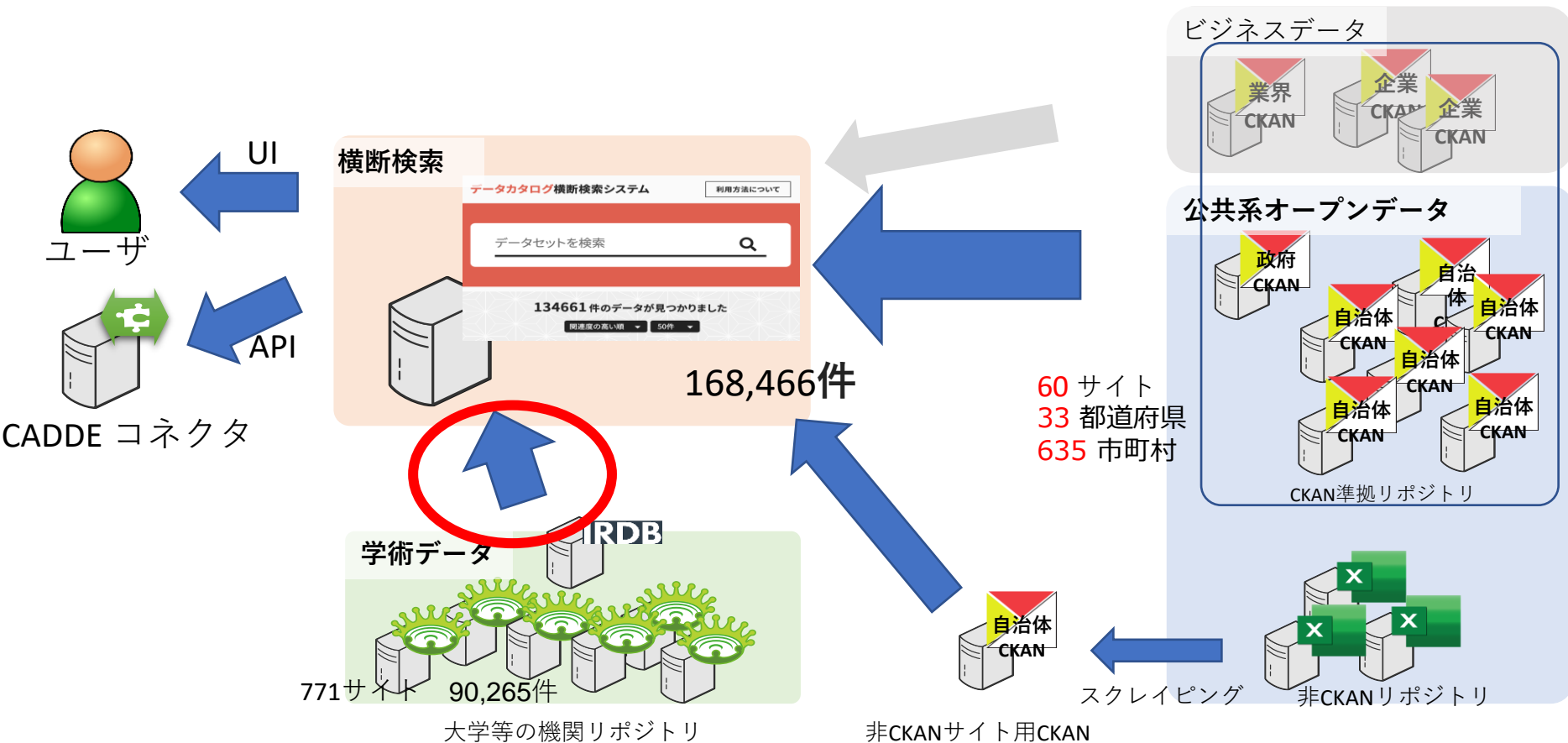
SIPデータカタログ項目データモデル
Dataset (データセット)
Distribution (配信)

CKAN データモデル
Package (パッケージ)
Resource (リソース)

★DCATのカタログクラスに対応するCKANの変数が存在しないため、CKANのConfigurationファイルにマッピングすることで対応

データカタログ横断検索

- ✓ 組織（企業、団体、自治体、大学など）が公開するデータカタログサイトからカタログ情報を収集して、検索可能にする。
- ✓ ユーザによるWeb UIを通じた簡易検索/詳細検索、CADDEコネクタによるAPI利用



60 サイト
33 都道府県
635 市町村

数字は2023年3月1日現在

<https://search.ckan.jp>

JPCOAR - CKAN mapping

root=/jpcoar:jpcoar/

CKAN Metadata	JPCOAR	Note
organization		Always use "irdb"
name	/oai:identifier	
title	/dc:title	
author	/jpcoar:creator/jpcoar:creatorName	Get an array.
notes	/datacite:description	Get an array and merge it into a string.
	/jpcoar:subject	Get an array and merge it into a string, then append to the description.
url	/jpcoar:identifier[@identifierType="URI"]	
metadata_modified	/datacite:date[@dateType="Issued"]	
metadata_created	/datacite:date[@dateType="Created"]	
site_name		Always use "学術機関リポジトリ".
resources[0].url	/jpcoar:URI	
resources[0].format	/jpcoar:mimeType	If not found, use "HTML".
resources[0].name		Always use "Fulltext".

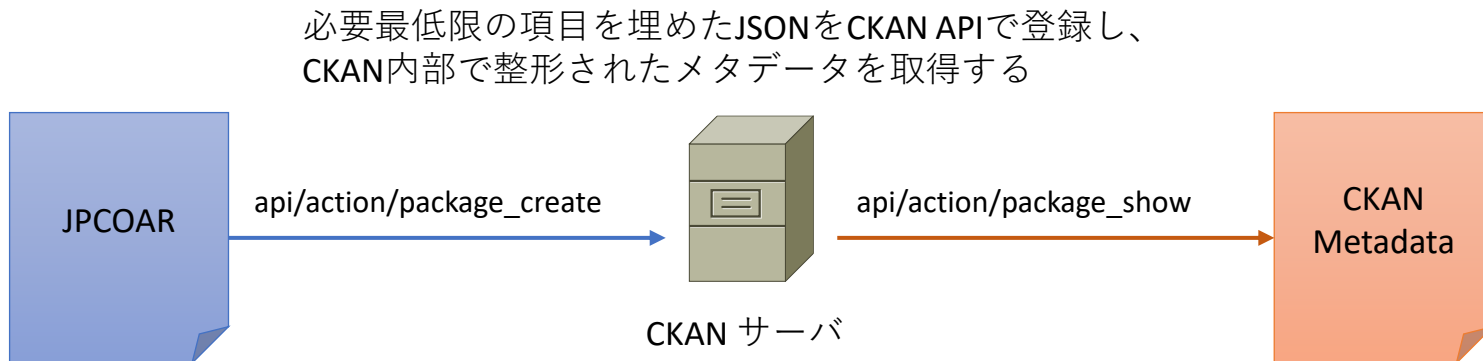
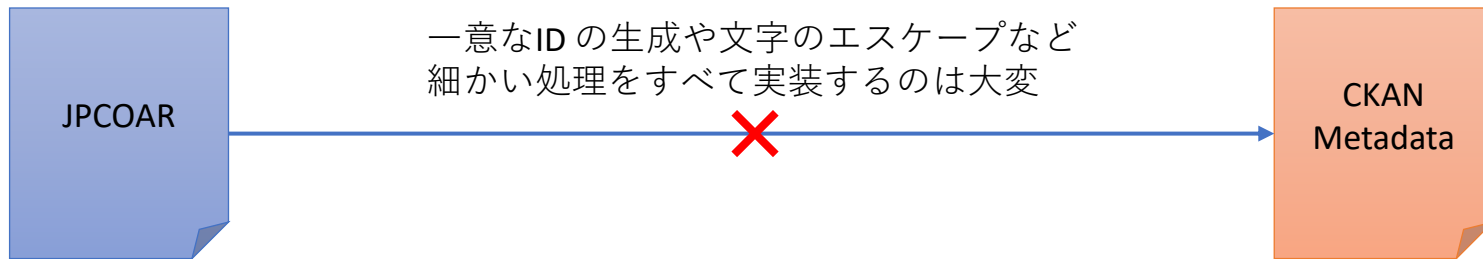
JPCOAR - CKAN mapping

root=/jpcoar:jpcoar/

CKAN Metadata	JPCOAR	Note
oai-identifier	/oai:identifier	
title_lang *1	/dcterms:alternative	Detect <i>lang</i> from "xml:lang" attribute.
extras.publisher_name	/dc:publisher	
extras.nii-type	/dc:type	
extras.language	/dc:language	
extras.sourceTitle	/jpcoar:sourceTitle	
extras.volume	/jpcoar:volume	
extras.issue	/jpcoar:issue	
extras.pageStart	/jpcoar:pageStart	
extras.pageEnd	/jpcoar:pageEnd	

*1 "title_lang" は title_ja, title_en のように記載されている言語の数だけ展開して生成する。

JPCOAR-CKAN Mapping: 転送プロセス



18.178.205.230__dataset:oai-irdb-nii-ac-jp-0:
2021

COVID-19/SARS-CoV-2関連のプレプリントに関する分散表現データ トー2020年05月17日版ー

<http://hdl.handle.net/11035/00006644> > dataset
「[学術機関リポジトリ](#)」に移動する

データセットについて

(no description)【リソース】Fulltext

データセット

[JSONダウンロード](#)

xckan_id	18.178.205.230__dataset:oai-irdb-nii-ac-jp-03078-0004530356
xckan_title	COVID-19/SARS-CoV-2関連のプレプリントに関する分散表現データセットー2020年05月17
xckan_site_name	学術機関リポジトリ
xckan_site_url	http://hdl.handle.net/11035/00006644
xckan_last_updated	2021-09-29T07:45:54Z
xckan_description	(no description)【リソース】Fulltext
xckan_tags	
データセットのライセンス	
maintainer	
relationships_as_object	
private	false
maintainer_email	
num_tags	0
id	3a4e0bc9-368f-460e-bfc7-e4f50688363f
データセットの発行日 (自動設定)	2021-09-29T07:45:54.382284

データセットの更新日または修正日 (自動設定)	2021-09-29T07:45:54.382293
author	小柴, 等, 林, 和弘
author_email	
state	active
version	
creator_user_id	23d469ef-c960-453f-854a-79898438210c
type	dataset
num_resources	1
license_id	
relationships_as_subject	
name	oai-irdb-nii-ac-jp-03078-0004530356
isopen	false
データセットの説明ページURL	
データセットの説明	(no description)
owner_org	35e0fa77-578c-4f18-ba47-9ecf3d92932a
データセットのタイトル	COVID-19/SARS-CoV-2関連のプレプリントに關
revision_id	3f10b9b9-d511-46d4-a7c3-002121b0ecbf
score	6.0150065

組織

description	
created	2021-09-16T08:11:45.405025
title	IRDB
name	irdb
is_organization	true
state	active
image_url	
revision_id	a239f758-aed1-46a4-b477-d1fc6e835d27
type	organization
id	35e0fa77-578c-4f18-ba47-9ecf3d92932a

<https://search.ckan.jp>

まとめ

- 現状

- DCATはデータカタログのメタデータのデファクト
- CKANはオープンデータポータルへのデファクト
- CKANはDCATにマッピングが可能
- JPCOAR->CKANはやってみた

- 課題

- CKANのメタデータは任意性が高い
 - 値のない項目も多い
- DCATはより構造化を志向しているが、どこまで構造化をするかは要検討
- CKAN->JPCOARもやっていますが、まだ道なれば