デジタル社会推進実践ガイドブック DS-469

メタデータ導入実践ガイドブック

2022年（令和4年）3月31日

デジタル庁

**〔キーワード〕**

メタデータ、DCAT-GOJ

**〔概要〕**

データ社会では様々なデジタルデータが活用、連携されるようになる。そのデジタルデータを見つけやすく、選定しやすくするために、データには見出し情報であるメタデータが付けられる。このメタデータは、各種コンテンツ、オープンデータ、ベース・レジストリなどで共通化されることで価値を増大させることができる。

本ペーパーではメタデータ導入や活用の考え方を示すとともに、その導入や活用のための方法を説明する。

改定履歴

|  |  |  |
| --- | --- | --- |
| 改定年月日 | 改定箇所 | 改定内容 |
| 2022年3月31日 | - | 初版決定 |

目次

[１ 背景と目的 1](#_Toc99390100)

[１.１ 背景 1](#_Toc99390101)

[１.２ 課題 2](#_Toc99390102)

[１.３ 目的 2](#_Toc99390103)

[１.４ メタデータとは何か 2](#_Toc99390104)

[２ メタデータ項目の定義 3](#_Toc99390105)

[２.１ DCATの概要 3](#_Toc99390106)

[２.２ DCATシリーズ 6](#_Toc99390107)

[1) BregDCAT 6](#_Toc99390108)

[2) GeoDCAT 6](#_Toc99390109)

[3) StatDCAT 7](#_Toc99390110)

[２.３ ベース・レジストリのメタデータ項目 7](#_Toc99390111)

[1) カタログ 8](#_Toc99390112)

[2) カタログレコード 10](#_Toc99390113)

[3) データセット 11](#_Toc99390114)

[4) データサービス 15](#_Toc99390115)

[5) ディストリビューション 16](#_Toc99390116)

[２.４ メタデータの記載例 19](#_Toc99390117)

[３ 様々な分野での活用 22](#_Toc99390118)

[３.１ ベース・レジストリでの検討 22](#_Toc99390119)

[1) カタログやデータセットの単位に関する検討 23](#_Toc99390120)

[2) データ項目のコード化、詳細化 25](#_Toc99390121)

[３.２ 教育分野での検討 25](#_Toc99390122)

[1) 既存メタデータとの比較 26](#_Toc99390123)

[2) 共通的なタグの必要性 26](#_Toc99390124)

[３.３ 会議資料での検討 27](#_Toc99390125)

[３.４ イベントや施設情報との共通化 28](#_Toc99390126)

[３.５ CKANへの実装の検討 28](#_Toc99390127)

[４ メタデータ付与による効果 29](#_Toc99390128)

[５ 解説 30](#_Toc99390129)

[５.１ データ標準 30](#_Toc99390130)

[1) W3C　DCAT 2.0 30](#_Toc99390131)

[2) EC　DCAT-AP v2.0.1 30](#_Toc99390132)

[3) EC　BregDCAT-AP v2.00 30](#_Toc99390133)

[4) Schema.org 30](#_Toc99390134)

[5) Google検索　データセット 30](#_Toc99390135)

# 背景と目的

## 背景

行政機関は国内最大のデータホルダーであるが、そのデータは行政内部でも社会全体でも十分に使いこなせているわけではない。そもそも、各行政機関がどのようなデータを保有し、どこまで活用可能なのかということは、その組織外からわからないだけでなく、同じ組織内でもわかっていないことは多い。

世界各国では、行政機関の保有するデータを管理する必要性に早くから気がつき、W3Cの定義するデータカタログのメタデータの標準であるDCAT(Data Catalog Vocabulary) [[1]](#footnote-1)を中心にデータ管理の取組を進めている。

また、従来のコンテンツの集積拠点としては図書館があり、既に体系的にコンテンツを管理している。他にも、教育、地理空間等の分野独自のメタデータを整備している分野もあり、イベントデータや施設等のデータも体系化が図られつつある。

これらのメタデータの基本構造は類似しており、相互運用性を確保することも可能と考えられている。

それに対応するため、令和２年12月25日に閣議決定された「デジタル・ガバメント実行計画」において、行政手続のワンスオンリーを実現するなど社会全体の効率性の向上を図るとともに、スマートシティなどの新しいサービスの創出を図るため、ベース・レジストリ[[2]](#footnote-2)の整備が喫緊の課題として挙げられています。また、「ベース・レジストリ・ロードマップ[[3]](#footnote-3)」では、ベース・レジストリで管理するデータを行政機関等が検索、利用できるようにするため、レジストリ・カタログを構築することとしています。また、レジストリ・カタログの検索性や管理を高度化するため、ベース・レジストリに登録するデータセットにはメタデータ[[4]](#footnote-4)を付与することが求められています。

これまでもオープンデータカタログや国立国会図書館の納本用メタデータ等のメタデータの活用事例はありましたが、今後は、一般のコンテンツ、オープンデータ、ベース・レジストリ等をシームレスにつなげるメタデータ管理が求められてきます。

## 課題

現在、国内の行政機関でメタデータの標準が存在しないため、データホルダー毎に自由にメタデータを管理しており、また、メタデータ自体が存在しない場合もある。このようにメタデータの活用が統一されていない状況となるため、データ利用者、データを管理する行政職員の双方に負担がかかるとともに、データを効率的に活用することができていない。

利用者が感じる課題

・データを見つけられない

・データの存在があるかどうかもわからない

・関連するデータに何があるのかわからない

・データを見つけてもライセンスなどの利用条件がわからない

・データの更新周期がわからない

・データの品質がわからない

行政職員が感じる課題

・他の行政機関が持っている情報を活用して業務を高度化、効率化したい

・データに関する問い合わせへの対応に負荷がかかる

## 目的

データセットに付与するメタデータは、ベース・レジストリやオープンデータ、コンテンツマーケットのカタログとの間でシームレスな検索が可能な仕組みとするため、国際標準に準拠することが重要になります。

本ペーパーは、メタデータの国際的な標準であるDCATに準拠したメタデータ項目を示すことで、コンテンツの発見性を高め、そのデータ連携や活用を容易にすること目的とします。

コンテンツにメタデータを付けることにより、カタログサイトだけではなく、データの取引市場、検索サイト、個々のコンテンツの利用等、様々な場面でデータが使いやすくなると考えられます。

## メタデータとは何か

メタデータとは、データに関するデータであり、データを検索する際などに使用します。データ関連データとしてデータに付帯して整備され、データカタログで検索に使用されます。対象地域、期間や品質情報などを含むことで、今後のデータ取引などで必須の情報になります。



図 1　メタデータのイメージ

# メタデータ項目の定義

## DCATの概要

DCATとは、国際標準化団体W3C(World Wide Web Consortium)が勧告しているデータカタログのメタデータ標準であり、その構造は図１のようなクラスとプロパティで示されています。

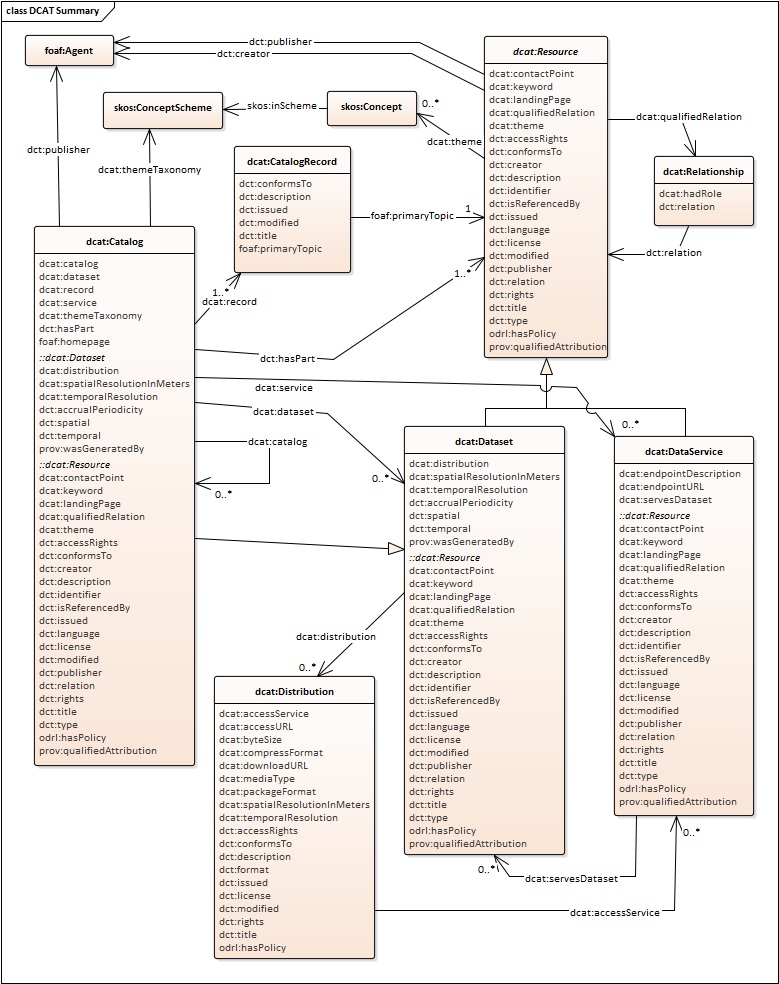


図 2　DCATモデルの概要

(出典：<https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>)

図２を概略化すると、以下のようにカタログ、カタログレコード、データセット、データサービス、ディストリビューションの階層構造で示すことができます。



図 3　DCATの構成概略

　また、上記概略図で示す各階層（クラス）の定義は以下のとおりです。

|  |  |
| --- | --- |
| カタログ | 特定目的に対するデータの集合体のことをカタログと呼ぶ。カタログは、より上位のカタログを持つことや、カタログ内にサブのカタログを持つこともある。 |
| カタログレコード | カタログの中で分野を設定したいときに持つ。カタログ冊子の章のようなものである。 |
| リソース | カタログの中のデータに関する項目である。データ項目は持たず、後述するデータセットやデータサービスにより構成される。 |
| データセット | 公開される情報の束であり、複数のデータの集合体であることもある。品質やライセンス等、利用するための情報を含む。 |
| データサービス | サブスクリプションやAPIでデータを入手するための詳細情報である。サービス形式などの情報を含む。 |
| ディストリビューション（配信） | 個々のデータ単位の情報である。配信するための技術的内容などを含む。 |

例えば、図２の構成を実際のデータセットに当てはめた場合、次の図３のように示すことができます。

例１：国土数値情報の場合



例２：会議体の場合



図 4　DCATの使用例

## DCATシリーズ

DCATには様々な派生形があります。EU（欧州連合）のEuropean Commission[[5]](#footnote-5)では、DCATの仕様に基づき、対象別にそれぞれ以下のようなメタデータ標準を定めています。

### BregDCAT

ベース・レジストリのメタデータ標準。行政機関の連絡先窓口やカタログの来歴情報などのメタデータ項目を定義[[6]](#footnote-6)しています。

### GeoDCAT

地理空間情報のメタデータ標準。住所や地理座標などのメタデータ項目を定義[[7]](#footnote-7)しています。

### StatDCAT

統計情報のメタデータ標準。統計データの系列や単位などのメタデータ項目を定義[[8]](#footnote-8)しています。

## ベース・レジストリのメタデータ項目

ベース・レジストリのメタデータ項目は、２．２で記したBregDCATを参考に、DCATの仕様に基づき、それぞれのクラスで定めるべき項目を「DCAT-GoJ」という名称で定義します。DCAT-GoJのイメージは図５のとおりです。

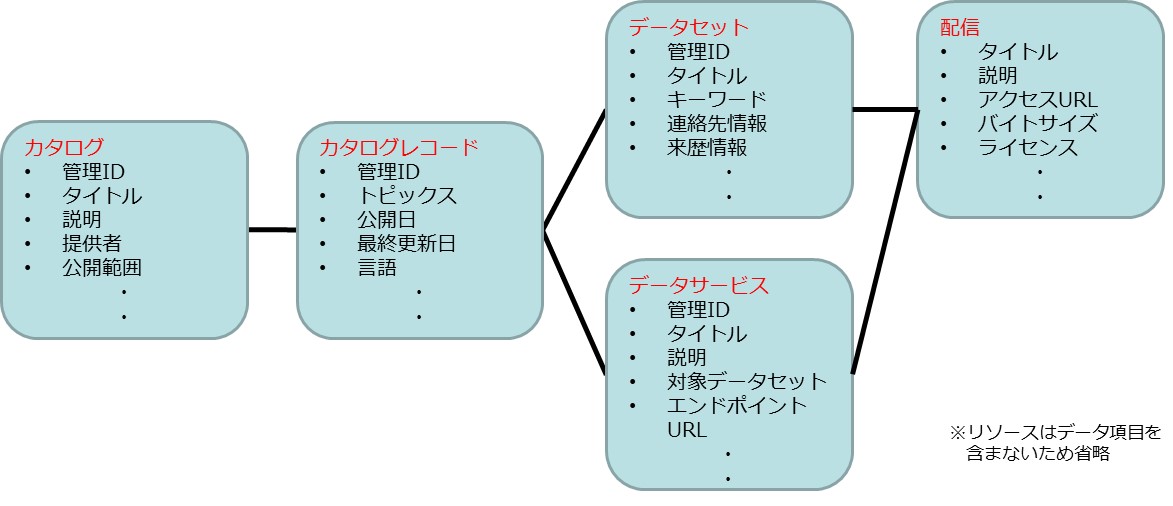


図 5　DCAT-GoJのイメージ

DCAT-GoJで定めるメタデータ項目は次に示すとおりです。

なお、次に示す表の内容についても以下のとおりとなります。

・メタデータ項目：メタデータの項目名を示しています。

・分類　　　　　：メタデータを付与するレベルを示しています。

　　　　　　　　　◎はメタデータの付与を必須とする項目、

○はメタデータの付与を推奨する項目、

空白はメタデータの付与を任意とする項目。

・項目説明　　　：メタデータ項目の説明を示しています。

・データ形式　　：メタデータのデータ形式を示しています。

・記述方式　　　：メタデータの記述方式を示しています。

　　　　　　　　　指定された形式で記述とされている場合は、項目説明欄で指定された形式で記述してください。

　　　　　　　　　指定された選択肢から選択とされている場合は、項目説明欄で示されている選択肢から選択してください。

### カタログ

| メタデータ項目 | 分類 | 項目説明 | データ形式 | 記述方式 |
| --- | --- | --- | --- | --- |
| 管理ID | ◎ | カタログをユニークに識別するための管理IDを示す。 | 文字列型 | 自由記述 |
| タイトル | ◎ | カタログのタイトルを示す。 | 文字列型 | 自由記述 |
| サブタイトル |  | カタログのサブタイトルを示す。 | 文字列型 | 自由記述 |
| 説明 | ◎ | カタログに収録されている情報の特徴を第三者に理解してもらうための説明を示す。また、関係性のあるカタログが存在する場合は、その名称も示す。 | 文字列型 | 自由記述 |
| キーワード | ○ | カタログを容易に検索できるように、検索タグとして扱うキーワードを示す。 | 文字列型 | 自由記述 |
| テーマ分類 | ○ | カタログが扱う情報のテーマ分類を示す。 | 文字列型 | 自由記述 |
| 対象地域 | ○ | カタログの対象地域を示す。 | 文字列型 | 自由記述 |
| 親カタログID | 〇 | 上位のカタログがある時にそのカタログIDを示す。 | 文字列型 | 自由記述 |
| 提供者 | ◎ | カタログを提供するエンティティ（組織又は個人）の名称を示す。  ※法人番号も検討 | 文字列型 | 自由記述 |
| 公開日 | ○ | カタログの公開日を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 最終更新日 | ○ | カタログを最後に更新・修正した日付を示す。 入力形式は以下のとおり。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 更新頻度 | ○ | カタログが更新される頻度を示す。 定期的に提供する場合、年・月・週・日あたりの回数を示す。 不定期の場合は、不定期と示す。 | 文字列型 | 自由記述 |
| 言語 | ○ | カタログがどの言語で記述されているかを示す。 なお、国際標準化機構（ISO）が発行する「ISO 639-1:2002」又は「ISO 639-2:1998」に準拠したアルファベット２文字又は３文字で表記する。 【例】日本語：ja、英語：en | 言語型 | 指定された形式で記述 |
| 公開範囲 |  | カタログの公開範囲を、以下のいずれかで示す。 ・公開 ・制限付き公開 ・非公開 | 文字列型 | 指定された選択肢から選択 |
| 公開条件 |  | カタログの公開範囲が「制限付き公開」の場合に、制限条件及び制限解除するための条件を示す。 | 文字列型 | 自由記述 |
| ライセンス | ○ | カタログに適用されるライセンスについて示す。 外部で定義されたライセンスへのリンク（URL）を示してもよい。 | 文字列型 | 自由記述 |
| 利用規約 |  | カタログの利用規約を示す。 外部で定義された利用規約へのリンク（URL）を示してもよい。 | 文字列型 | 自由記述 |
| ホームページ | ○ | カタログに関連するホームページを示す。 | 文字列型 | 自由記述 |

### カタログレコード

| メタデータ項目 | 分類 | 項目説明 | データ形式 | 記述方式 |
| --- | --- | --- | --- | --- |
| トピックス | ◎ | カタログのトピックスを示す。 | 文字列型 | 自由記述 |
| 説明 | ◎ | カタログレコードにどのようなデータセットが収録されているか、説明を示す。 | 文字列型 | 自由記述 |
| 公開日 | ○ | カタログレコードの公開した日付を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 最終更新日 | ○ | カタログレコードを最後に更新・修正した日付を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 言語 | ○ | カタログレコードがどの言語で記述されているかを示す。 なお、国際標準化機構（ISO）が発行する「ISO 639-1:2002」又は「ISO 639-2:1998」に準拠したアルファベット２文字又は３文字で表記する。 【例】日本語：ja、英語：en | 言語型 | 指定された形式で記述 |

### データセット

| メタデータ項目 | 分類 | 項目説明 | データ形式 | 記述方式 |
| --- | --- | --- | --- | --- |
| 管理ID | ◎ | データセットをユニークに識別するための管理IDを示す。 | 文字列型 | 自由記述 |
| タイトル | ◎ | データセットのタイトルを示す。 | 文字列型 | 自由記述 |
| サブタイトル |  | データセットのサブタイトルを示す。 | 文字列型 | 自由記述 |
| バージョン |  | データセットのバージョン情報を示す。 | 文字列型 | 自由記述 |
| 説明 | ◎ | データセットの特徴を第三者に理解してもらうための説明を示す。 | 文字列型 | 自由記述 |
| キーワード | ○ | データセットを容易に検索できるように、検索タグとして扱うキーワードを示す。 | 文字列型 | 自由記述 |
| 対象地域 | ○ | データセットの対象地域を示す。 | 文字列型 | 自由記述 |
| 対象期間 | ○ | データセットの対象期間を示す。 入力形式は以下のとおりとする。 ただし、明確に年月日で示せない場合は、自由記述とする。 ・開始年月日/終了年月日 : YYYY-MM-DD/YYYY-MM-DD | 文字列型 | 指定された形式で記述 または 自由記述 |
| 分類 | ○ | データセットが扱うデータの分類を日本標準産業分類に基づき示す。 | 文字列型 | 指定された形式で記述 |
| 提供者 | ○ | データセットを提供するエンティティ（組織又は個人）の名称を示す。  ※法人番号も検討 | 文字列型 | 自由記述 |
| 作成者 |  | データセットの作成に関わったエンティティ（組織又は個人）を示す。  ※法人番号も検討 | 文字列型 | 自由記述 |
| 連絡先情報 | ○ | データセットについて問い合わせを行う際の連絡先情報として以下の項目を示す。 ・組織名 ・部署名 ・電話番号 ・メールアドレス ・フォームURL  ※法人番号も検討 | 文字列型 | 指定された形式で記述 |
| タイプ |  | データセットの形式を示す。 | 文字列型 | 自由記述 |
| 来歴情報 |  | データセットの出所や変更履歴等の来歴情報を示す | 言語型 | 自由記述 |
| 品質評価 |  | データセットの品質評価（評価方法は別途定義）を示す。 | 文字列型 | 自由記述 |
| 品質測定結果 |  | データセットの品質測定結果（測定方法は別途定義）を示す。 | 文字列型 | 自由記述 |
| 公開日 |  | データセットを公開した日付を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 最終更新日 | ○ | データセットを最後に更新・修正した日付を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 更新頻度 |  | データセットが更新される頻度を示す。 定期的に提供する場合、年・月・週・日あたりの回数を示す。 不定期の場合は、不定期と示す。 | 文字列型 | 自由記述 |
| 言語 |  | データセットがどの言語で記述されているかを記載する。 なお、国際標準化機構（ISO）が発行する「ISO 639-1:2002」又は「ISO 639-2:1998」に準拠したアルファベット２文字又は３文字で表記する。 【例】日本語：ja、英語：en | 言語型 | 指定された形式で記述 |
| 公開範囲 |  | データセットの公開範囲を、以下のいずれかで示す。 ・公開 ・制限付き公開 ・非公開 | 文字列型 | 指定された選択肢から選択 |
| 公開条件 |  | カタログの公開範囲が「制限付き公開」の場合に、制限条件及び制限解除するための条件を示す。 | 文字列型 | 自由記述 |
| 準拠する標準 |  | データセットが準拠する標準を示す。 | 文字列型 | 自由記述 |
| 関連ドキュメント |  | データセットに関する情報を持つドキュメントへのリンク（URL）を示す。 | 文字列型 | 自由記述 |
| ランディングページ |  | データセットに関して、追加・補足できる情報が公開されているWebページのURLを示す。 | 文字列型 | 自由記述 |

### データサービス

| メタデータ項目 | 分類 | 項目説明 | データ形式 | 記述方式 |
| --- | --- | --- | --- | --- |
| 管理ID | ◎ | データサービスをユニークに識別するための管理IDを示す。 | 文字列型 | 自由記述 |
| タイトル | ◎ | データサービスのタイトルを示す。 | 文字列型 | 自由記述 |
| 説明 | ◎ | データサービスの特徴を第三者に理解してもらうための説明を示す。 | 文字列型 | 自由記述 |
| キーワード | ○ | データサービスを容易に検索できるように、検索タグとして扱うキーワードを示す。 | 文字列型 | 自由記述 |
| 対象データセット | ○ | データサービスが配信対象とするデータセットを示す。 | 文字列型 | 自由記述 |
| 提供者 | ○ | データサービスを提供するエンティティ（組織又は個人）の名称を示す。 | 文字列型 | 自由記述 |
| タイプ |  | データサービスの形式を示す。 | 文字列型 | 自由記述 |
| 公開範囲 |  | データサービスの公開範囲を、以下のいずれかで示す。 ・公開 ・制限付き公開 ・非公開 | 文字列型 | 指定された選択肢から選択 |
| 公開条件 |  | カタログの公開範囲が「制限付き公開」の場合に、制限条件及び制限解除するための条件を示す。 | 文字列型 | 自由記述 |
| ライセンス |  | データサービスを利用可能にするためのライセンスを示す。 外部で定義されたライセンスへのリンク（URL）を示してもよい。 | 文字列型 | 自由記述 |
| 準拠する標準 |  | データサービスが準拠する標準を示す。 | 文字列型 | 自由記述 |
| 関連ドキュメント |  | データサービスに関する情報を持つドキュメントへのリンク（URL）を示す。 | 文字列型 | 自由記述 |
| エンドポイントURL | ○ | データサービスのルートの位置又は主要エンドポイント(Webで解決可能なIRI(Internationalized Resource Identifier))を示す。 | 文字列型 | 自由記述 |
| ランディングページ |  | データサービスに関して、追加・補足できる情報が公開されているWebページのURLを示す。 | 文字列型 | 自由記述 |

### ディストリビューション

| メタデータ項目 | 分類 | 項目説明 | データ形式 | 記述方式 |
| --- | --- | --- | --- | --- |
| タイトル | ◎ | ディストリビューションのタイトルを示す。 | 文字列型 | 自由記述 |
| 説明 | ◎ | ディストリビューションに関する説明を示す。 | 文字列型 | 自由記述 |
| アクセスサービス |  | ディストリビューションにアクセスできるサービスを示す。 | 文字列型 | 自由記述 |
| バイトサイズ |  | ディストリビューションのサイズをバイト単位で示す。 | 数値型 | 指定された形式で記述 |
| 圧縮形式 |  | ディストリビューションのデータサイズを縮小するために圧縮されている場合、圧縮形式を示す。 | 文字列型 | 自由記述 |
| メディアタイプ |  | ディストリビューションのメディアタイプを示す。 IANA(Internet Assigned Numbers Authority)管理語彙で定義されている配信データのメディアタイプから選択する。 | 文字列型 | 指定された形式で記述 |
| 公開日 | ○ | ディストリビューションを公開した日付を示す。 入力形式は以下のとおりとする。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 最終更新日 | ◎ | ディストリビューションを最後に更新・修正した日付を示す。 入力形式は以下のとおり。 ・YYYY-MM-DD（西暦-月-日） | 日付型 | 指定された形式で記述 |
| 期間 | ○ | ディストリビューションの配信期間を示す。 入力形式は以下のとおりとする。 配信期間 : YYYY-MM-DD/YYYY-MM-DD | 日付型 | 指定された形式で記述 |
| ステータス |  | ディストリビューションのステータスを以下のいずれかで示す。 ・配信中 ・配信準備中 ・配信停止中 | 文字列型 | 指定された選択肢から選択 |
| 言語 |  | ディストリビューションデータがどの言語で記述されているかを記載する。 なお、国際標準化機構（ISO）が発行する「ISO 639-1:2002」又は「ISO 639-2:1998」に準拠したアルファベット２文字又は３文字で表記する。 【例】日本語：ja、英語：en | 言語型 | 指定された形式で記述 |
| ライセンス | ○ | ディストリビューションを利用可能にするためのライセンスを示す。 外部で定義されたライセンスへのリンク（URL）を示してもよい。 | 文字列型 | 自由記述 |
| 利用規約 |  | ディストリビューションの利用規約を示す。 外部で定義された利用規約へのリンク（URL）を示してもよい。 | 文字列型 | 自由記述 |
| 準拠する標準 |  | ディストリビューションが準拠する標準を示す。 | 文字列型 | 自由記述 |
| 関連ドキュメント |  | ディストリビューションに関する情報を持つドキュメントへのリンク（URL）を示す。 | 文字列型 | 自由記述 |
| アクセスURL | ◎ | ディストリビューションを取得できるWebページのURLを示す。 | 文字列型 | 自由記述 |
| ダウンロードURL |  | ディストリビューションがダウンロード可能なURL を示す。 | 文字列型 | 自由記述 |

## メタデータの記載例

前記で示したDCATに基づきメタデータ情報を記載した例は次のとおりです。

| 記載例 | | |
| --- | --- | --- |
| カタログ  管理元で別途付番するため、記載不要です。 | | |
|  | 管理ID | CA00001  分類が必須(◎)でなく、未定または不明な項目は空欄可です。 |
|  | タイトル | 文字情報基盤 |
|  | サブタイトル |  |
|  | 説明 | 行政で用いられる人名漢字、法人用漢字、かな等の約６万文字の漢字です。 |
|  | キーワード | 漢字、戸籍統一文字、住民基本台帳ネットワーク統一文字、登記統一文字、変体仮名 |
|  | テーマ分類 | 文字 |
|  | 対象地域 | 全国  指定された形式で記述してください。 |
|  | 提供者 | 文字情報技術促進協議会 |
|  | 公開日 | 2011-10-26 |
|  | 最終更新日 | 2020-8-26 |
|  | 更新頻度 | 不定期  指定された選択肢から選択してください。 |
|  | 言語 | ja |
|  | 公開範囲 | 公開 |
|  | 公開条件 |  |
|  | ライセンス | IPAフォントライセンスv1.0 |
|  | 利用規約 | クリエイティブ・コモンズ 表示 – 継承 2.1 日本 ライセンス条件 |
|  | ホームページ | https://moji.or.jp/mojikiban/ |
| カタログレコード | | |
|  | トピックス | 文字 |
|  | 説明 | 行政で用いられる人名漢字、法人用漢字、かな等の約６万文字の漢字です。 |
|  | 公開日 | 2011-10-26 |
|  | 最終更新日 | 2020-8-26 |
|  | 言語 | ja  管理元で別途付番するため、記載不要です。 |
| データセット | | |
|  | 管理ID | CA00001-DST001 |
|  | タイトル | MJ文字情報一覧表 |
|  | サブタイトル |  |
|  | バージョン | Ver.006.01 |
|  | 説明 | 各文字に関するコード、読み、字母、画数等をまとめた情報。 |
|  | キーワード | フォント、ヨミガナ、画数 |
|  | 対象地域 | 全国 |
|  | 対象期間 |  |
|  | 分類 | 全ての業務 |
|  | 提供者 | 文字情報技術促進協議会 |
|  | 作成者 | 情報処理推進機構 |
|  | 連絡先情報 | 組織名：一般社団法人 文字情報技術促進協議会 メールアドレス：info@moji.or.jp フォームURL：https://moji.or.jp/about/contact/ |
|  | タイプ | Strict Open XML |
|  | 来歴情報 | 2020年10月に、情報処理推進機構から文字情報技術促進協議会に信託譲渡 |
|  | 品質評価 | 正確性、網羅性 |
|  | 品質測定結果 | 公務で使うのに十分な品質 |
|  | 公開日 | 2011-10-26  別途定義する基準に基づき記載してください。 |
|  | 最終更新日 | 2020-8-26 |
|  | 更新頻度 | 不定期 |
|  | 言語 | ja |
|  | 公開範囲 | 公開 |
|  | 公開条件 |  |
|  | 準拠する標準 |  |
|  | 関連ドキュメント |  |
|  | ランディングページ |  |
| データサービス | | |
|  | 管理ID | DSV00001 |
|  | タイトル | 文字情報基盤検索サービス  管理元で別途付番するため、記載不要です。 |
|  | 説明 |  |
|  | キーワード |  |
|  | 対象データセット |  |
|  | 提供者 |  |
|  | タイプ | web |
|  | 公開範囲 | 公開 |
|  | 公開条件 |  |
|  | ライセンス |  |
|  | 準拠する標準 |  |
|  | 関連ドキュメント |  |
|  | エンドポイントURL |  |
|  | ランディングページ |  |
| 配信 | | |
|  | タイトル | mji.00601-xlsx.zip |
|  | 説明 |  |
|  | アクセスサービス |  |
|  | バイトサイズ | 7.3M |
|  | 圧縮形式 | zip |
|  | メディアタイプ |  |
|  | 公開日 |  |
|  | 最終更新日 | 2019-05-01 |
|  | 期間 |  |
|  | ステータス | 配信中 |
|  | 言語 | ja |
|  | ライセンス |  |
|  | 利用規約 |  |
|  | 準拠する標準 |  |
|  | 関連ドキュメント |  |
|  | アクセスURL | <https://moji.or.jp/mojikiban/mjlist/> |
|  | ダウンロードURL | <https://moji.or.jp/wp-content/mojikiban/oscdl/mji.00601-xlsx.zip> |

# 様々な分野での活用

広範な対象でメタデータが共通的に使えるように、様々な分野で検証を行った。

## ベース・レジストリでの検討

国は今後ベース・レジストリを整備して行く予定である。日本のベース・レジストリは複雑な構造を持っており、法人の情報では、商業登記情報に法人名と本店所在地があるが、国税庁で法人番号を付番するとともに、漢字情報を一般のデジタル機器で使える範囲に代替するなどのクレンジング処理を行い公開している。また、その他のベース・レジストリ情報を統合して、経済産業省のgBizInfo[[9]](#footnote-9)が情報の公開を行っている。このように関係性が分かりやすいものを明確にカタログ化して行くことが必要であり、そのためにはメタデータが重要である。また、メタデータでデータに関する情報をわかりやすく提供する際には、品質情報を含む利用を検討するために重要な情報も明確にして行く必要がある。

そのため、ベース・レジストリのカタログ化のためのメタデータが必要となり、欧州のベース・レジストリ用のメタデータであるBregDCAT[[10]](#footnote-10)をベースに検討を行った。

### カタログやデータセットの単位に関する検討

例えば、医療機関のベース・レジストリを例として表現すると、厚生労働省が医療機関の一覧として各地方局単位で公開している「保険医療機関・保険薬局の指定等一覧及び保険医・保険薬剤師の新規登録一覧」から、以下のように表現される。

グラフィカル ユーザー インターフェイス, アプリケーション, メール

自動的に生成された説明

図 6　医療機関の一覧

このページを、１つのデータセットととらえた場合は、以下のように、データを表現することができる。

カタログ ：保険医療機関・保険薬局の指定等一覧及び保険医・保険薬剤師の新規登録一覧

データセット ：関東甲信越厚生局管内

ディストリビューション ：茨城県医科（PDF）

：茨城県医科（歯科併設）（PDF）

：茨城県歯科（PDF）

：茨城県歯科（医科併設）（PDF）

：茨城県薬局（PDF）

：栃木県医科（PDF）

：・・・

しかし、このようにするとディストリビューションが多過ぎて検索性が低くなってしまう。また全国の都道府県単位で活用したいこともあることから、都道府県単位に以下のように表現することもできる。

カタログ ：保険医療機関・保険薬局の指定等一覧及び保険医・保険薬剤師の新規登録一覧

カタログレコード ：関東甲信越厚生局管内

データセット ：茨城県

ディストリビューション ：医科（PDF）

：医科（歯科併設）（PDF）

：歯科（PDF）

：歯科（医科併設）（PDF）

：薬局（PDF）

また、医科、歯科、薬局のようにカテゴリで表現することも可能である。

カタログ ：保険医療機関・保険薬局の指定等一覧及び保険医・保険薬剤師の新規登録一覧

カタログレコード ：関東甲信越厚生局管内

データセット ：医科

ディストリビューション ：茨城県医科（PDF）

：茨城県医科（歯科併設）（PDF）

：栃木県医科（PDF）

：栃木県医科（歯科併設）（PDF）

：・・・

カタログレコードを使わずに、以下のようにカタログを親子形式でネスト化[[11]](#footnote-11)して表現することも可能である。

親カタログ ：保険医療機関・保険薬局の指定等一覧及び保険医・保険薬剤師の新規登録一覧

カタログ ：関東甲信越厚生局管内

データセット ：茨城県

ディストリビューション ：医科（PDF）

：医科（歯科併設）（PDF）

：歯科（PDF）

：歯科（医科併設）（PDF）

：薬局（PDF）

このように、カタログは目的に応じて編集が可能で、様々な対象の記述のために自由度を持ち、かつ、検索の利便性も維持した形式にする必要がある。

### データ項目のコード化、詳細化

DCATをベースに項目の検討をして行くと、項目の記述方法に困ることがある。

例えば、公開者（Publisher）という項目は、Agentというデータ項目で表現することとなっている。つまり、厚生労働省がPublisherの場合、Agentとしての詳細情報として「名称」「法人番号」「所在地」等の情報を持つことができるが、どこまでデータ項目を持つべきかと行った問題がある。また、DCATの品質情報は自由記述になっている。しかし、自由記述ではデータ利用に関する品質評価を客観的に行うことが困難であり、さらなる詳細化が求められている。

このようなデータ項目に関して実装のための記述方法の検討が必要である。

## 教育分野での検討

教育分野では、教材のメタデータであるIEEE LOM[[12]](#footnote-12)やコースウェアのメタデータであるIMSのCC (Thin Common Cartridge)[[13]](#footnote-13)が存在する。このように、既存のメタデータがある場合には、その整合性が重要となる。

### 既存メタデータとの比較

既存のメタデータであるIEEE LOMとDCAT、及び、以前文部科学省が運用を行っていた教育コンテンツデータベースであるNICERのメタデータ[[14]](#footnote-14)との比較を試みた。以下はIEEE LOMとDCATの比較の簡易整理である。



図 7　メタデータ比較の例

保有するコンテンツにDCATのメタデータがついていたとしても、LOMをベースとした教育システムでは必要な情報が不足することがある。また、連絡先などのデータ項目の構成や表記がLOMとDCATの実装で異なる場合もある。その場合には、データのコンバータやタグ追加の仕組みなどが必要となってくる。

### 共通的なタグの必要性

教材には学習指導要領コードの付与が推進されているが、一般のコンテンツに学習指導要領コードは付与されていない。そこで汎用的なコードとの連携が重要になる。

国内で広く普及しているコードが望ましく、その点からは、統計を中心に使用される日本標準産業分類と図書館の分類に使用される日本十進分類法であるNDC[[15]](#footnote-15)の汎用性が高いと考えられる。また、グローバルなデータ流通を考えても、このような汎用性の高いコード体系を参照しておくことが重要になる。



図 8　共通的なタグ活用のイメージ

## 会議資料での検討

行政機関の会議資料は、国内外の動向をまとめているなど価値のあるデータであるが、報告書のみしか参照されないことが多い。また、インターネットで検索すると会議資料が検索結果として提示されるが、そこにメタデータは付与されていない。特に、「資料１」と記載されているが、その会議体が何かわからないということが多く、メタデータの付与が求められている（昨今の行政サイトはディレクトリ構造ではなくパーマリンク[[16]](#footnote-16)が使われることが多いため、関連情報を探すことが難しくなっている）。

会議資料のメタデータは、以下のように表現できる。

カタログ ：○○会議

カタログレコード ：△△タスクフォース

データセット ：第一回○○会議△△タスクフォース

ディストリビューション ：資料１

：資料２

：参考資料１

：データ

：議事録

また、カタログをネスト構造にして会議資料のメタデータを以下のように表現することもできる。

親カタログ ：○○会議

カタログ ：△△タスクフォース

データセット ：第一回○○会議△△タスクフォース

ディストリビューション ：資料１

：資料２

：参考資料１

：データ

：議事録

どちらの表記も可能であるが、利用者の利便性を考えると、会議資料に関して統一的なルールを作ることが望ましい。

## イベントや施設情報との共通化

メタデータはコンテンツに付けられることが多いが、利用者の立場から考えると、「図 8　共通的なタグ活用のイメージ」のように施設やイベント情報と共通的に検索することができることでさらに利便性が向上する。タイトルや概要、タグ等のデータ定義を共通化しておくことで、このような検索も実現可能となる。

## CKAN[[17]](#footnote-17)への実装の検討

データカタログサイトを作る際に、データカタログのオープンソースであるCKANを使用することが多い。しかし、CKANは完全にDCATに準拠しているわけではない[[18]](#footnote-18)。実装にあたっては、DCATエクステンションを使用するが、個々のデータ定義やコード設計含め、メタデータの設計は他のサービスとの連携も含め精査して行く必要がある。

# メタデータ付与による効果

メタデータ付与の検討を行うと、現場の負荷が増えるため協力することができないという反対を受けることが多い。その作成する瞬間だけを考えると多少作業は増えるが、１時間もかからない簡単な作業である。

一方、メタデータがないと利用者はそのデータを検索するために長い時間を浪費することになる。それでも見つけることができればよい方であり、必要なデータにたどり着けず、そのデータを使用した効果を得られないことも多い。通常は、メタデータを付ける機会は一度きりであるが、利用する機会は複数回に及ぶことが多い。その全体の効果を作成者も考える必要がある。

メタデータ作成者にとっても効果がある。効果を以下に示す。

・過去の情報検索が容易になる

・人事異動の際に引継ぎが容易になる

・問い合わせに対して迅速に対応することができる

・データの利用が増え、政策的な効果が上がる

・データカタログなどを用いて広く広報することができる

それでも、メタデータを付けるのは負荷であると考える部門もある。メタデータは、納品物の一部として納品してもらう仕組みにして、そのデータをアップロードするだけにするといった工夫も可能である。

自分にとってのメリット、利用者にとってのメリット、データの価値最大化の視点から、メタデータの付与を推進して行くことが重要である。

# 解説

## データ標準

### W3C　DCAT 2.0

インターネットに関する標準化団体のW3Cが2020年４月に公開したデータカタログ用メタデータであり、現在、バージョン３の検討が進められている。

<https://www.w3.org/TR/vocab-dcat-2/>

### EC　DCAT-AP v2.0.1

DCAT 2.0をベースにEC諸国で導入するために実務的に編集した実装モデルである。

<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/201-0>

### EC　BregDCAT-AP v2.00

欧州委員会が、DCATをベースにベース・レジストリのカタログ用に作成したメタデータである。

<https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap>

### Schema.org

Web検索のデファクトスタンダードのデータモデルを策定しているSchema.org が提供するデータセットのメタデータモデルである。

<https://schema.org/Dataset>

### Google検索　データセット

Googleの検索で使用されるデータセットのメタデータである。

<https://developers.google.com/search/docs/data-types/dataset?hl=ja>

1. <https://www.w3.org/TR/vocab-dcat-2/> [↑](#footnote-ref-1)
2. ベース・レジストリとは、公的機関等で登録・公開され、様々な場面で参照される、人、法人、土地、建物、資格等の社会の基本データであり、正確性や最新性が確保された社会の基盤となるデータベース群の総称。 [↑](#footnote-ref-2)
3. <https://www.kantei.go.jp/jp/singi/it2/dgov/dai10/siryou_a.pdf> [↑](#footnote-ref-3)
4. データに関するデータと呼ばれる見出し情報である。データ検索や管理に使われる。 [↑](#footnote-ref-4)
5. 欧州連合の政策執行機関。 [↑](#footnote-ref-5)
6. 詳細は<https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap>を参照。 [↑](#footnote-ref-6)
7. 詳細は<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/geodcat-application-profile-data-portals-europe>を参照。 [↑](#footnote-ref-7)
8. 詳細は<https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/statdcat-application-profile-data-portals-europe>を参照。 [↑](#footnote-ref-8)
9. <https://info.gbiz.go.jp/> [↑](#footnote-ref-9)
10. <https://joinup.ec.europa.eu/collection/access-base-registries/solution/abr-bregdcat-ap> [↑](#footnote-ref-10)
11. あるカタログがある際に、カタログを束ねた上位のカタログがある場合や、サブ分類による下位のカタログがある場合がある。 [↑](#footnote-ref-11)
12. <https://www.ieeeltsc.org/working-groups/wg12LOM/lomDescription/> [↑](#footnote-ref-12)
13. <https://www.imsglobal.org/activity/common-cartridge#:~:text=Thin%20Common%20Cartridge%20%28Thin%20CC%29%20is%20a%20profile,CCs%20only%20contain%3A%20Learning%20Tools%20Interoperability%20%C2%AE%20Links> [↑](#footnote-ref-13)
14. <https://nier.jp/papers/6-8.pdf> [↑](#footnote-ref-14)
15. <https://www.ndl.go.jp/jp/data/NDC10code201708.pdf> [↑](#footnote-ref-15)
16. 資料のURLを、乱数等を使って一意に特定する仕組みである。URLがサイトの構造と独立しているため、サイトの更新があってもコンテンツのリンク切れを防ぐことができる。 [↑](#footnote-ref-16)
17. <https://github.com/ckan> [↑](#footnote-ref-17)
18. <https://github.com/ckan/ckanext-dcat#rdf-dcat-to-ckan-dataset-mapping> [↑](#footnote-ref-18)