# 3 地理空間情報と LOD

応用

松澤有三（インディゴ（株））

筆者らは 2014 年からリンクト・オープン・データ（Linked Open Data，LOD）のための地理識別子基盤である「GeoNames.jp」を運営している．本稿ではその経験から得られた知見をもとに，地理座標による場所表現と地理識別子による場所表現の特徴，地理識別子基盤の実例を紹介し，地理識別子の相互連携，他分野への応用について解説する．

## 地理空間情報と場所の表現手法

2007 年に制定された地理空間情報活用推進基本法を契機に，行政機関による地理空間情報のオープン化が推進されてきた．かつては専門分野であった地理情報システム（Geographic Information System，GIS）も Web 地図などを通じて広く一般に普及している．今日ではオープンデータの利活用の手法として，地図上でさまざまな情報を重ね合わせて可視化するアプリケーションは一般化してきている．

さて，地理空間情報と LOD の関係を考えるにあたっては，データと場所の関係について掘り下げることが必要である．GIS 分野では，データと場所を関連付けることを空間参照と呼んでおり，大きく分けて 2 つの方法が定義されている．1 つは緯度経度などの座標を用いることで，空間上の場所を直接参照する方式である．もう 1 つは住所・郵便番号・区域コードといった，座標によらない場所表現を使用して，空間上の場所を間接的に参照する方式である．以下，これらの場所表現手法の特徴と LOD との関係を紹介する．

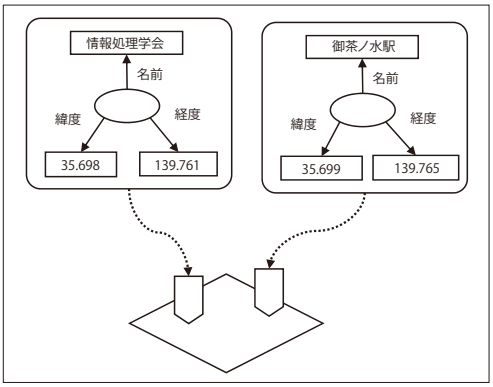


図 -1　座標による位置表現例

### ● 地理座標による場所表現

オープンデータの典型的な利用例として，「緯度経度の付与されたデータを地図上でマッシュアップ」というものがある（図 -1）．また，緯度経度の付与されていないデータに対して，住所などをヒントに緯度経度を付与するジオコーディングという手法もよく用いられている．

このような座標を中心とした地理空間情報は，オープンデータにおける機械可読データの重要性，緯度経度座標という共通の座標を用いてデータを整備することのメリットなど，オープンデータの推進における数々の教訓を残しているし，今後もその有効性は変わらないだろう．

しかし，このような座標による地理空間情報が外部のデータと連携する LOD であるかは注意して観察する必要がある．緯度経度座標によってデータと座標系の関係が整備されてはいるものの，特別に意識しない限りはデータとデータの連携はなされない．データに座標や適切な語彙が付与されても，外部へのリンクがなく閉じたデータセットは案外多いのではないだろうか．

### ● 地理識別子による場所表現

他方，地理識別子を用いた地理空間情報において

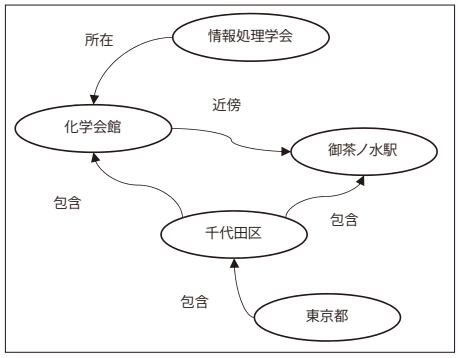


図 -2　地理識別子による位置表現例

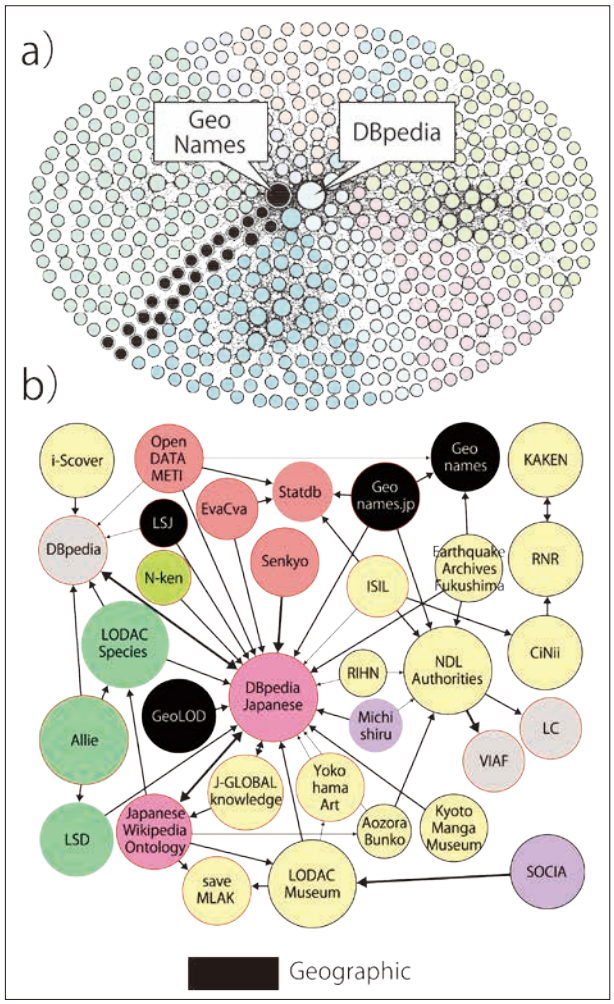


図 -3　LOD クラウド図における地理空間情報
（a）2014 年世界版 [1]　（b）2015 年 11 月日本版 [2]

は，地理識別子と地理識別子の関係を表現することで場所が表現されるという特徴がある（図 -2）．このようなデータモデルは LOD の考え方と親和性が高い．

図 -3-（a）は 2014 年版の LOD クラウド図 であり，世界の LOD のデータセットの規模，分類，関係性が表現されている．図中の左下に位置する黒い円群が地理情報として分類されているデータセット群である．その中の GeoNames.org は地理識別子の代表的なデータセットであり，地理情報分野に限らず，多様なデータセットからリンクされている．また，データセットとともに地理識別子を使用する語彙の整備も進んでいる．

図 -3-（b）は 2015 年版の日本の LOD クラウドであるが，GeoNames.org を参照するデータセットは 3 つであり，日本国内ではまだ地理識別子を用いたデータ連携は黎明期にあるといえる．

### ● LOD としての地理空間情報

日本では地理座標によるデータ整備や可視化といった手法はかなり一般化してきているが，今後 LOD としての地理空間情報の整備を推進するにあたっては，地理識別子の利用促進が欠かせない．

次章からは地理識別子として利用可能なデータセット，地理識別子の相互運用性を担保するリンクセット，それらの応用について解説していく．

## 地理識別子のデータセット

地理識別子のデータセットは，GeoNames のようにもともと地理識別子の整備を目的としたものもあるし，DBpedia のようにデータセットの中に地理識別子として利用できるものが含まれているものもある．日本版 LOD クラウドに登場しているデータセットから日本国内の地理識別子を収録しているものを紹介する（表 -1）．

### ● 例 1：GeoNames.org

GeoNames.org は世界中の地名に対して URI を与える地名データセットである．Web ページ上では 800 万以上の地名の提供を謳っている．各地名に対して数字からなる GeoNames ID を付与することで地名を識別できるようにし，これを用いた URI を整備するという仕組みになっている．

各国の地名集や郵便番号データなどをもとに基本のデータベースが作成され，利用者が Web・ユー

| データセット | 地理識別子 URI（千代田区の例） | 特徴 |
|---|---|---|
| GeoNames.org | http://sws.geonames.org/1864529/ | 世界の地名に独自 GeoNames ID を付与 |
| GeoNames.jp | http://geonames.jp/resource/ 東京都千代田区 | 日本の地名を都道府県名から記述することで URI の一部としている |
| Statdb | http://data.e-stat.go.jp/lod/sac/C13101-19830818<br>http://data.e-stat.go.jp/lod/sac/C13101-19700401<br>http://data.e-stat.go.jp/lod/sac/C13101 | 標準地域コードと更新時期によって構成される URI，および最新版を指す URI を提供 |
| DBpedia Japanese | http://ja.dbpedia.org/resource/ 千代田区 | Wikipedia の記事名に由来する URI |
| NDL Authorities | http://id.ndl.go.jp/auth/ndlna/00306437 | 地名の典拠データも整備されている |
| GeoLOD | http://geolod.ex.nii.ac.jp/resource/51sMnl | GeoNLP の地名辞書を LOD 化 |

表 -1　地理識別子データセットの例


図 -4　GeoNames.jp

ザインタフェース を通じて地名を修正・追加することもできる．

LOD 分野における地理識別子の第一選択肢ともいえるデータセットであるが，実際に日本の地名を検索してみると，市区町村以下の地名の整備状況が網羅的でない／日本語表記が誤っている／日本語表記がない地名がある，といった課題があり，日本の LOD からのリンク先としては扱いづらい状況である．

## ● 例 2：GeoNames.jp

筆者らが開発運用にあたっている GeoNames.jp は GeoNames.org の日本ローカル版を目指して作成された，日本の地名に日本語の URI を与える地名データセットである（図 -4）．都道府県郡市区町村および，その直下の町名・字・丁目までを収録対象としており，約 36 万の地名を収録している．「http://geonames.jp/resource/」のあとに都道府県から始まる地名を記述したものを URI として使用できるため，住所文字列の一部から地理識別子を得るような場合に利便性が高い．

データソースとしては国土数値情報・行政区域データおよび街区レベル位置参照情報（いずれも国土交通省，おおむね年に 1 回の更新がある）を加工して作成しており，過去から現在に至る地名の提供を目指している．

## ● 例 3：都道府県・市区町村コード情報

都道府県および市区町村の区域を示すコードとして，1970 年から標準地域コードが整備されてきた．これは都道府県市区町村に対して 5 桁の数字からなるコードを付与するもので，合併や区域の変更に応じて都度改正されてきた．JIS X0401，JIS X0402 として標準化もされている．

この標準地域コードを LOD 形式で整備したものが総務省統計局から提供されている．表 -1 に例示されているように，合併や区域の変更に応じて URI が更新されていくという特徴がある．すでに標準地域コードを付与した形で作成されたデータに対して地理識別子を付与する場合には，本データセットの利便性は高い．

なお，都道府県・市区町村コード情報は 2013 年 12 月より総務省統計局・次世代統計利用システムの一部として試行的に提供されてきたが，2016 年 3 月より総務省統計局 e-Stat LOD サイトの一部として正式提供が開始された．移行に伴って語彙およびリソース URI のドメイン部分が変更されているために，今後のデータの作成にあたっては，正式版のリソース URI を参照するのが望ましい．

## ● 例 4：DBpedia

DBpedia は Wikipedia から抽出した情報を LOD として公開するコミュニティプロジェクトである．Wikipedia には地名や自治体に関する記事も多数含

まれているために，このような記事に対応する DBpedia のリソースは地理識別子として扱うことも可能である．

Wikipedia の特性上，都道府県や市区町村といった自治体については対応するページが整備されているが，市区町村配下の地名の整備状況は一様ではない．小地域の地理識別子が必要な場合には，GeoNames.jp のようなほかの選択肢を検討することも必要となってくる．

通常は「http:/ja.dbpedia.org/resource/」のあとに都道府県市区町村名を付与することで地理識別子として使用できる URL が得られるが，「府中市」のように同名の自治体が複数存在する場合には http://ja.dbpedia.org/resource/ 府中市 _（東京都）のような曖昧さのない URL を使用する必要がある．

### ● 例 5：Web NDL Authorities

「国立国会図書館典拠データ検索・提供サービス（Web NDL Authorities）」は，国立国会図書館が維持管理する典拠データを RDF 形式で整備・提供している．典拠データの中には都道府県市区町村名や古地名が含まれており，この種の URI は地理識別子の一種として使用が可能である．

ほかの地理識別子データセットが地名・行政区域を直接モデリングしているのに対して，典拠データは書誌情報を分類管理することを目的に作成されており，シソーラスや分類表を表現するための語彙である SKOS を用いてモデリングされているという特徴がある．

### ● 例 6：GeoLOD

国立情報学研究所 GeoNLP 開発チーム[☆1] では，地名情報処理システム GeoNLP で使用する地名辞書を LOD 対応したものを GeoLOD として公開している．地名辞書に収録されているデータは政府系オープンデータを中心に，市区町村・大字・鉄道駅・空港・河川・山など多岐にわたる．

☆1 http://agora.ex.nii.ac.jp/GeoNLP/

## 地理識別子のリンクセット

地理識別子を含むデータセットが多数存在し，また，相互に重複するデータが整備されていることが分かるが，データ作成者はどの地理識別子にデータをリンクさせるべきだろうか？　あるいは複数の地理識別子にリンクしなければならないのだろうか？異なるデータセットに含まれる地理識別子間の同一性を整備共有することによって，特定の地理識別子にリンクしたデータが間接的にほかの地理識別子にリンクするような環境が整う．ここではこのような相互運用性を確保するための「リンクセット」の仕組みと実例を紹介する．

### ● データセットとリンクセット

データセットのメタデータを表現する語彙を定義する VoID[☆2] では，「データセット」と「リンクセット」という用語を定義している．

DBpedia や GeoNames.org といった，ある目的に沿って単一のコミュニティ・ドメインによって整備されるデータの集積を「データセット」と呼ぶのに対して，異なるデータセットのリソース間の関係記述に特化したデータセットのことを「リンクセット」と呼んで区別している．

多様なコミュニティから多様なデータセットが提供される昨今においては，地理識別子に限らず，このようなリンクセット整備の必要性は増してくるだろう．

### ● リンクセットの語彙

地理識別子に限らず，リンクセットの整備にあたっては，主語となるリソースと目的語となるリソースの関係を適切に表現するための語彙の選定が重要である．特にデータセットの提供者の意図を損ねるような語彙の使用は避けなければならない．

owl:sameAs は主語リソースと目的語リソースが完全に同一であることを表現する語彙である．VoID の仕様書では GeoNames.org と DBpedia のリ

☆2 http://www.w3.org/TR/void/

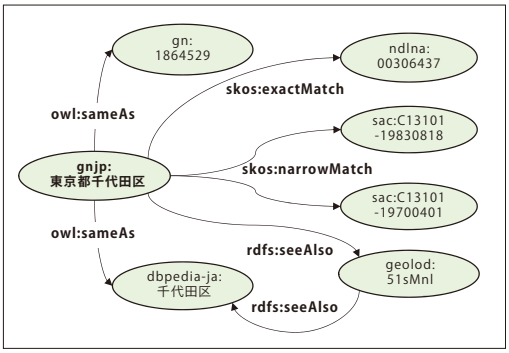


図 -5　地理識別子間のリンク例

ソースを owl:sameAs でリンクさせる例題が取り上げられるなど，リンクセットの文脈で目にすることも多い語彙である．しかし，主語リソースと目的語リソースが「完全に同一」と言い切れない場合には owl:sameAs は適切ではない．

rdfs:seeAlso は主語リソースに関する追加情報が目的語リソースから得られることを指す，関連性の表現として使用される．

skos:Concept のインスタンスとして記述されたリソース間の関係を表現するための語彙としては，skos:mappingRelation から派生する closeMatch, exactMatch, broadMatch, narrowMatch, relatedMatch の各語彙が検討の対象となるだろう．これらは SKOS において異なる概念体系に所属する概念間の関係をマッピングするために用意されている．

### ● 地理識別子のリンクセット

表 -1 に掲載した千代田区の地理識別子について，GeoNames.jp の千代田区を中心に各リソースとの関係をグラフにしたものが図 -5 である．このようなリンク関係を整備することによって，たとえば GeoNames.org の URI を起点として国立国会図書館の SPARQL Endpoint から情報を得る，といった利用方法の基礎ができる．ここでは前節の各種語彙を選択して関係を付与しているが，特に owl:sameAs の使用には注意が必要である．標準地域コードのように，期間ごとに別々に定義されたリソースが同一の扱いとなってしまうような作用があるためである．

筆者らは地理識別子の基盤整備の一環として GeoNames.jp を起点とした各種地理識別子データセットへのリンクセットを整備し，GitHub 上で公開を行っている．本稿執筆時点では，図 -3-（b）における国内主要データセットのうち，DBpedia Japanese，GeoNames.org，Web NDL Authorities，Statdb（都道府県・市区町村コード情報），および GeoLOD に対して都道府県郡市区町村の同一性に基づくリンクセットを公開している．GeoNames.jp の維持管理と並行して，これらのリンクセットの管理を継続しつつ，連携先の開拓，過去の地名や小地域を含む整備範囲の拡大を模索している．

## 地理識別子の応用

本章では地理識別子を参照する語彙，データの事例として，Web NDL Authorities で想定されているような文書・資料データの分類，GeoNames.org で想定されているような地理情識別子間のリンク，標準地域コードで想定されているような統計分野への応用について紹介する．

### ● 文書・資料・データの分類

データやドキュメントのメタデータを記述するために，Dublin Core ボキャブラリは広く使用されている．データ分類を記述するための語彙である dcterms:subject を地理識別子と組み合わせることで，データやドキュメントが当該地域に関する情報であることを明示できる（図 -6-（a））．

CKAN のようなデータカタログサイトにおいて，市町村ごとのデータを抽出するような使い方へも応用ができるだろう．

### ● 地理情報のリンク

地理情報と地理情報の位置関係を扱うための語彙は各種存在するが，ここでは応用範囲の広い関係として包含関係の事例を紹介する（図 -6-（b））．

包含関係を表現するための語彙として，GeoNames Ontology では parentFeature が，Schema.org では containsPlace および containedInPlace が定義されている．

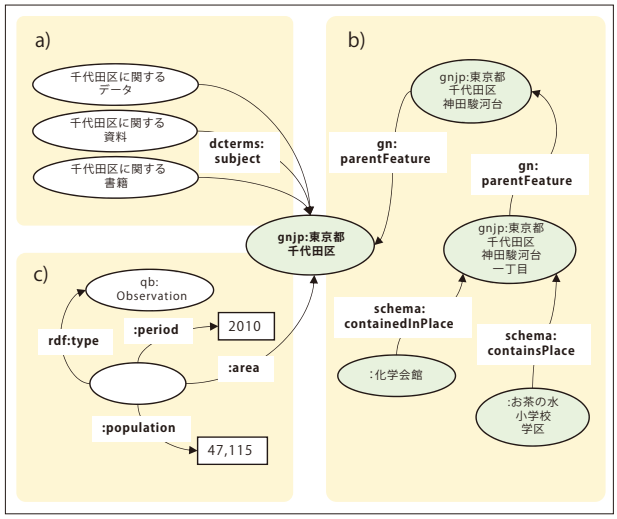


図 -6　地理識別子の応用例

GeoNames.org や GeoNames.jp といった基盤では parentFeature を用いた階層情報が基本情報として整備されており，ここに containedInPlace を用いて自データをリンクすることで，地理的な所属関係を表すことができる．いままで何物ともリンクしていなかった地理空間情報をリンクさせるための初手としておすすめできる方法である．

また選挙区や学区といった区域の表現にも包含関係は活用できる．従来はこのようなデータ構造は選挙区や学区内の住所文字列の羅列，地図への図示，といったかたちでなければ記述が難しかったが，地理識別子と包含関係を用いることで，機械可読性の高いかたちで表現することを可能にしている．

ここで挙げた包含関係以外にも，近傍・隣接，さらには幾何的な交差関係やネットワーク的な接続関係など，多様な応用が期待される．

### ● 統計分野への応用

行政のオープンデータでは統計に関するデータも豊富に提供されている．統計の LOD 化については 2014 年の W3C The RDF Data Cube Vocabulary の勧告によって語彙の整備が一段落し，実運用の段階に入っている．総務省統計局は 2015 年より統計データの LOD 化として Data Cube を使用したデータ提供を検討中である．

図 -6-（c）は「2010 年の千代田区の人口」を Data Cube のデータモデルで表現した例である．観測データを意味する qb:Observation のインスタンスから，地域・時間・人口を参照することで統計データが表現される仕組みである．ここで，地域を特定するために地理識別子が使用されている．

統計データが対象とする地域は国・都道府県・市区町村・小地域やメッシュ，さらにこれらの組合せで表現される地方・地域など多岐にわたる．今後の統計データの LOD 化の流れの中で，地理識別子の有効利用が期待される．

## 今後の展望

本稿では地理空間情報の LOD 化における地理識別子の意義，応用を中心に解説した．実際に使用可能な地理識別子も登場しており，地理識別子にリンクするデータの作成を通じて，地理空間情報の LOD 化が推進されていくものと期待している．語彙の適切な使用方法の模索や有益なアプリケーションの開拓は今後の継続的な課題であろう．

筆者らの開発運用する GeoNames.jp は，今後もデータセットのメンテナンスを行い，ほかの地理識別子基盤と連携すべくリンクセットの整備を継続していく．

**参考文献**
1）Schmachtenberg, M., Bizer, C., Jentzsch, A. and Cyganiak, R. : Linking Open Data Cloud Diagram 2014, http://lod-cloud.net/
2）Kato, F. : 日本語 Linked Data Cloud 図 2015-11-18 版.
（2016 年 4 月 1 日受付）

**松澤有三** ■ yuzo@indigo.co.jp
東京大学工学系研究科修士課程修了，工学修士．2011 年よりインディゴ（株）シームレス空間基盤研究開発センター主席研究員．