

モビリティサービスを事例とした 複数ドメインデータ相互運用メカニズムの検討

藤田 幸久¹・松澤 有三²・高橋 陽一³・桑原 昌広⁴

¹非会員 トヨタ自動車株式会社 コネクティッド先行開発部 (〒100-0004 東京都千代田区大手町 1-6-1)

E-mail: yukihisa.fujita@toyota-tokyo.tech (Corresponding Author)

²正会員 インディゴ株式会社 シームレス空間基盤研究開発センター (〒154-0004 東京都世田谷区太子堂 4-1-1)

E-mail: yuzo@indigo.co.jp

³正会員 インディゴ株式会社 シームレス空間基盤研究開発センター (〒154-0004 東京都世田谷区太子堂 4-1-1)

E-mail: yoit@indigo.co.jp

⁴正会員 トヨタ自動車株式会社 コネクティッド先行開発部 (〒100-0004 東京都千代田区大手町 1-6-1)

E-mail: kuwahara@toyota-tokyo.tech

センシング技術の発達に伴い様々なデータの取得、流通が推進されている。一方、スマートシティのように対象となるデータのドメインや管理主体が多岐にわたる場合、データ形式や粒度の違いから横断的な利用が困難となる。そこで本研究では、異なるドメインおよび管理主体のデータを横断的に活用するために、既存標準やOSSを活用したデータ集約の実現と、第三者が使いやすい形で流通させる手法を提案する。提案手法では、ドメインごとのデータモデル定義およびドメイン横断で利用する共通マスタ定義を用意し、共通のエンドポイントから統一的に参照可能とすることで、データ変換の手間を低減しつつ使い勝手の良いデータ流通を実現する。スマートシティ上でのモビリティサービスを想定し、提案手法を用いたサンプルを実装した結果、利用者側からの容易なデータ流通が確認できた。

Key Words: *City OS, Data Interoperability, Smart City, Data Model*

1. はじめに

(1) 背景

ICT およびセンシング技術の発達により、様々な情報が計測されデータとして流通するようになってきている。例えば、CAN (Controller Area Network) データと呼ばれる車両のドアの開閉や速度、位置等によるデータや、ウェアラブルデバイス等を通じて収集される個人の心拍数や加速度、位置のデータ等、対象は多岐に渡る。データ活用の観点では取得可能なデータ数の増加は歓迎される一方、データの管理、流通の観点で新たな課題が生じる。一企業が提供するサービスであれば、自社でデータ定義やID体系を独自定義した上で流通させることが可能であった。しかし、複数のステークホルダーが参画する、例えばスマートシティのようなケースでは、各ステークホルダーが独自に保持するデータを収集し、横断的に利用可能なように関連付けた上で流通させる必要がある。

ここではスマートシティ内を運行するモビリティサービスを例に考える。モビリティサービスの運行にはその都市内の道路や路線の情報だけでなく、物理的、法的な規制による恒常的な通行可否、事故による一時的な

規制情報等といった情報が必要となる。これらの情報は、日本においては道路情報はOpenStreetMap¹等の地図データやDRM (Digital Road Map)²から、一時的な規制情報に関してはJARTIC³が提供しているため、各提供者から個別に取得する必要がある。これら提供元の異なるデータを取得後、利活用に向けて統合する際は以下のような問題が発生しうる。

問題 1. データ項目の不一致

問題 2. マスタ体系の不一致

問題 3. 利用プロトコルの不一致

データ項目の不一致は、各データが持つ各項目の命名や意味が提供元に異なる問題である。例えば、同じ道路の長さを持つデータ項目でも、“length”や“distance”等が考えられ、命名が同じだった場合でもキロメートルかメートルか等、意味合いが異なる場合がある。各データはデータの提供元が本来の目的に最も都合の良い形式で保持されるため、異なる目的のデータを統合する場合にこのような問題が生じやすい。マスタ体系に関しても同様であり、国道か私道の識別をする場合に、数字のIDで保持するか、自然言語で所持するか等、様々なケースが存在する。さらに、これらのデータは、

提供元が REST API での提供であったり、フォーマットが JSON や CSV 等、様々な形式で保持、提供されており、データを利用するためのプロトコルを統一する必要がある。統合が必要なデータが数種類かつ一時的であれば一定回数の変換のみで大きな問題にはならないが、データの種類の増加した場合や、継続的な変換が必要である場合、あるいはデータ形式が途中で変化する場合には都度変換する方式は非現実的になる。また、データの提供元が異なるドメイン、例えばナビデータと建築データを統合する場合、利用されるデータ形式の違いや利用目的の違いもあり、データ統合はさらに困難になる。

(2) 目的

本研究では、前述の課題に対応するため、複数のステークホルダーがそれぞれ個別に管理しているデータ群を横断的に活用する仕組みと必要な基盤を検討する。このような課題への対応として、都市 OS と呼ばれるスマートシティ向けプラットフォームの一つである FIWARE¹⁾ は Smart Data Models²⁾ としてスマートシティをはじめとした複数の分野で利用可能なデータモデルを定義している。Smart Data Model は実際のスマートシティや関連プロジェクトで実際に利用されたデータモデルを採用しており、現在も拡大を続けている。Smart Data Models では NGSI (Next Generation Service Interface)³⁾ およびそれを Linked Data 形式に拡張した NGSI-LD が採用されており、当該データモデル、形式を採用すればプラットフォーム間のデータ流通が容易になることが期待される。他方、現状においては都市内のデータを表現するために十分なモデルが定義されておらず、またマスタ体系も統一されていない。

そこで本研究では、単一のデータモデルに依存するのではなく、複数のデータモデルや標準を活用してデータ流通を促進する仕組みを提案する。提案する仕組みでは以下の 3 点を要件とする。

- 要件 1. ドメイン単位のデータモデル
- 要件 2. ドメイン横断で共通参照可能なマスタ体系
- 要件 3. 全ドメインに対する単一のエンドポイント

単一のデータモデルですべてを表現することはせず、一定のドメインという単位でデータモデルを規定した上で、ドメイン間のデータ流通、共通のマスタ体系を提供する。これにより、単一のデータモデルを強制的に適用するのではなく、既存のデファクト、デジュールを流用しつつも、利用者側での変換の手間を低減することを目指す。また、それらを単一のエンドポイントで提供することで、異なるデータモデルに起因するデータのサイロ化を防ぐ。本稿では、特に以下の 2 点につ

いて述べる。

- ドメイン単位でのデータ管理、流通のコンセプト提案およびサンプル実装
- 上記サンプル実装を用いた提案手法の評価

2. 関連研究

(1) データモデルを取り巻く状況

様々な情報を単一の仕組みで扱う方法として、前述したように単一のデータモデルで必要な事象を表現する方法が挙げられる。モビリティサービスが関係するデータモデルの例を表-1 に示す。例えば地理情報で言えば Shape や GML がそれぞれデファクト、デジュールであったり、モビリティでは GTFS をベースとして様々な派生が定義されている。これらのモデルは、特定のモデルが他のモデルの上位互換となっていることは無く、特定の目的にフォーカスされていることが多い。例えば GeoJSON は高さに関する情報を持つことはできるが、建築物の構造や設備などの設計情報の表現は不向きである。他方、IFC で表現される建築情報等は建築物の構造や設計を詳細に表現することに適しているが、地図上で他の地理空間情報を重畳し俯瞰的に分析する用途には不向きである。

目的ごとに定義されたデータモデルと異なり、多目的で活用可能を目指すモデルとして FIWARE の Smart Data Models や Schema.org が存在する。Smart Data Models は名前が表すように Smart X (City, Building 等) でのデータ流通促進を目的として、Schema.org はインターネット上のコンテンツに対する“語彙”として定義されている。一方で、多目的を目指す汎用的なモデルは特定の目的向けに特化したモデルと比較した場合に、活用の利便性やカバレッジの観点で不足しがちである。

上記のように、目的別や対象物によってそれぞれデータモデルが定義されており、複数の“標準”が乱立している状況と言える⁴⁾。一方で、本来ドメイン固有に存在する目的に沿ってデータモデルは定義されることが多いため、単一のデータモデルが多くのドメインをカバーするような状況は考えづらい。そこで、データモデルの統一や無理な適用は避け、複数のデータモデルの活用を考える必要がある。

(2) 相互運用を目指したプラットフォーム

いくつかの分野で分野内の相互運用を目指したプラットフォームが提案されている。例えば、日本の防災分野においては、各自治体のデータを統合、活用するための SIP4D¹⁷⁾ というデータ流通基盤が運用されている。SIP4D では、自治体がそれぞれ収集、管理している避難所の情報や被災情報等に対し、データ形式の変換、メッ

表-1 モビリティが関係する各ドメインごとの代表的なデータモデル

ドメイン	データモデル名
地理情報	GML ⁵⁾ , GeoJSON ⁶⁾ , Shape ⁷⁾
空間情報	GML ⁵⁾ , CityGML ⁸⁾ , IndoorGML ⁹⁾
建築物	IFC ¹⁰⁾ , RealEstateCore ¹¹⁾ , BRICK ¹²⁾
モビリティ	GTFS ¹³⁾ , GBFS ¹⁴⁾ , Transdatamodel ¹⁵⁾
多目的	Smart Data Models ²⁾ , Schema.org ¹⁶⁾

シュ付与等の共通処理を施した上で、ダッシュボード等からの一元的な利用を実現している。また、IoT のデータ流通に対しては、Fed4IoT プロジェクト⁴で作成された VirIoT¹⁸⁾ がデータ流通を実現するプラットフォームとして挙げられる。VirIoT では、多種多様な IoT から送信されたデータを集約し、NGSI-LD のフォーマットで流通させることが可能になる。

これらのプラットフォームは防災や IoT といった特定の用途、ドメインに特化することでデータ流通を実現している。一方で、本研究で課題としているようなドメイン間をまたいだデータ流通は対象としていない。本研究では、このような相互運用を目指したプラットフォームの考え方と前述のデータモデルを組み合わせ、ドメインごとには固有のデータモデルを、ドメイン間は別の相互運用の仕組みを用意する。

3. 提案手法

(1) 概要

提案する方式の概要を図-1 に示す。ドメインは単一のデータモデルを保持する最小単位とし、各データソースは何かしらのドメインに所属する。例えば、図中右側で定義された地理情報は道路や関連した情報を保持するドメインであり、それらを単一のデータモデル (e.g., NGSI-LD, EUROTL) で保管する。各ドメインは横断処理用の API と固有処理用の API をそれぞれ備える。ドメイン横断での利用とドメイン固有の処理を可能にする。また、各ドメイン間でのマスタの不一致が生じないように、共通マスタとして URI により定義されたデータ群を備える。

上記の仕組みにより前述の各要件を満たす機能を提供する。以下、各要件ごとに備える機能を詳述する。

(2) ドメイン単位のデータモデル

提案手法では、ドメイン単位で単一のデータモデルを定義し、当該データモデルに情報を集約する。一般的に、地理情報であれば空間演算、ビル等の建築物であれば構造計算等、ドメインごとに固有の処理が存在

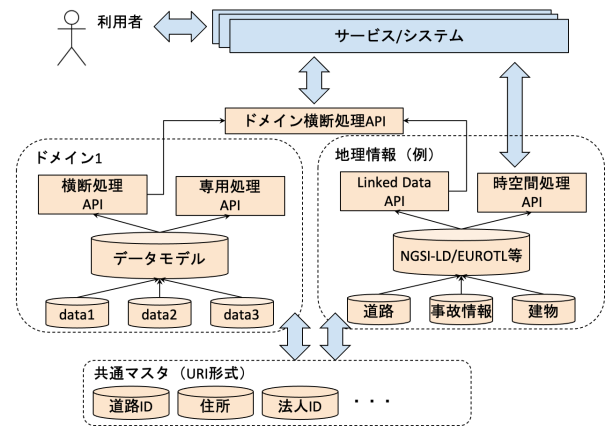


図-1 提案コンセプトの概要

する。ドメインをまたいだ互換性に偏重し、そのような固有の処理が高コストあるいは不可能になることは避ける必要がある。そこで、提案手法では、そのような固有の処理が必要な単位をドメインとし、当該ドメイン内ではデータモデルを統一する。これにより対象ドメイン内で個別定義されたデータ群の管理を容易にしつつ、ドメイン固有の処理を実施可能にする。ただし、データモデルの選定時には、単一のエンドポイントにて参照可能にできるよう、Linked Data として参照可能な形式 (JSON-LD 等) か、当該型式へ変換が可能なモデルのみ扱う。

(3) ドメイン横断で共通参照可能なマスタ体系

ドメイン単位でデータモデルが統一されたとしても、マスタが異なってしまえば各データが関連づいたとは言えない。例えば、道路の片道通行という事象を表現しようとした場合に、あるデータモデルでは事前定義されたコード (e.g., OW) で、他方では自然言語 (e.g., One Way) で表現される可能性がある。結果として、本来同様のことが表現され、また相互運用で同じ対象を抽出可能なはずなのに、データの解釈ができず同じ処理が不可能になる。少量のデータであれば、コードや表現の対応表を作成して対応可能であるが、量の増加、範囲の拡大に伴い管理が対応表の作成は指数関数的に必要な工数が増加する。

そこで、本提案では、すべてのデータモデルから同一のマスタ体系を URI にて参照する方式を採用する。同様の考え方として、公的機関が整備する信頼性の高い再利用可能な基礎情報であるベースレジストリが挙げられる (日本の IMI 語彙基盤⁵ や欧州の Joinup⁶ 等)。ただし、公的機関が整備するベースレジストリは都市や地域固有の詳細なコードやマスタといった点は含まれていない。そこで、本研究で扱う共通マスタは、URI 形式で参照可能な統一したコードやその表現を規定す

表-2 サンプル実装で利用したデータおよび変換先データモデル

データ名	取得元	元のデータ形式	変換先 (モデルクラス名)
道路	OpenStreetMap	OSM (XML)	Transportation.Road Transportation.RoadSegment Transportation.TransportStation
交通規制	JARTIC	CSV, PDF	Transportation.RestrictedTrafficArea
工事	国土交通省 近畿地方整備局 近畿地方道路情報提供システム	JSON	Transportation.CityWork
事故	警察庁交通事故統計情報	CSV	Transportation.RoadAccident
バス	兵庫県バス路線情報	GTFS	UrbanMobility 内の GTFS モデル群
天候	気象庁	JSON	PointOfInterest.PointOfInterest Weather.WeatherObserved

る。これらのコードはドメインの各データモデルに依存しない表現であり、必要に応じてデータモデル内の語彙との関係性を定義する。これにより、ドメインをまたいだ場合でも同一の対象物を同一の方法で表現可能になる。

(4) 全ドメインに対する単一のエンドポイント

ドメイン単位で個別のデータモデルを採用する場合、ドメインをまたいだ処理を実施する際に何かしらの仕組みまたは変換が必要となる。一方で、利用者側が各ドメインに個別の問い合わせをする必要がある場合、実質的にサイロ化したデータ群と変わらない。そこで、ドメインをまたいだ横断処理に対し、単一のエンドポイントで実施可能な仕組みを用意する必要がある。

このようなエンドポイントを提供する仕組みとしては、Linked Data に対するクエリ言語である SPARQL⁷ が挙げられる。SPARQL は Linked Data を表現するための RDF に対し、様々な処理を実施するクエリ言語だが、複数のエンドポイントに対する Federated Query が発行可能な仕組みを持つ⁸。そこで本提案では、各ドメインごとに Linked Data 形式のエンドポイントを定義した上で、当該エンドポイント群に対して Federated Query を発行する単一のエンドポイントを用意する。これにより、利用者側は内部に存在する複数のドメインを意識することなく、横断的な処理が可能になる。

4. データモデルに関するサンプル実装

(1) 概要

提案手法を実装した場合に生じる課題を把握するため、データモデル部分を中心にサンプルを実装した。サンプル実装では、提供元が異なる複数のデータソースをモビリティサービスで利用するユースケースを想定した。モビリティサービスを実現する場合、ある地点間

の移動に関する情報を正確に把握する必要がある。すなわち、ある地点から別の地点に移動する際の経路、およびその経路の利用可否、現状での規制等を取得する必要がある。他方、道路や交通規制といった情報は、様々なデータホルダーが独自のデータ形式で保持しているため、提案手法を用いて複数種類のデータを一元的に扱えるように管理する。

(2) 利用データ

表-2 に、本実装で利用したデータを示す。本実装ではモビリティサービスでの利用をユースケースとしているため、道路および交通規制、追加の規制となりうる工事、事故情報、天候、ほかサービスとなるバスの路線情報を対象データに含めた。なお、これらのデータは対象となる地域によって取得の量や粒度が異なる。本実装では、各データの量の観点から、2022年2月時点での兵庫県神戸市周辺のデータを利用した。

表-2 記載の通り、対象のデータは独自定義の XML や CSV, JSON で提供されており、共通的な処理を実施するためにはデータ変換が必要になる。本実装ではこれらのデータを、Smart Data Models で定義されたデータモデル⁹ に準じて NGSI-LD 形式に変換した。Smart Data Models はモデル内の各クラス（以降、モデルクラスと呼ぶ）で Transportation や UrbanMobility 等の表現があり、本ユースケースとの親和性が高い。一方で、本実装では NGSI-LD に準じた独自モデルクラスを定義することはせず、可能な限り Smart Data Models に定義されたモデルクラスのみを用いた。これは、NGSI-LD でデータ流通が確保されたとしても、独自モデルクラスが存在することでデータのかけあわせや比較が困難になることを避けるためである。

各データに対する変換先のモデルクラスは、単一のモデルクラスの場合もあれば複数のモデルクラスの組み合わせの場合もありうる。例えば、道路情報を変換し

```
// Transportation.RoadSegment
{
  "@context": [
    "https://smartdatamodels.org/context.
      jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-
      ld-core-context.jsonld"
  ],
  "type": "RoadSegment",
  "id": "urn:ngsi-ld:RoadSegment
    :675980797",
  "location": {
    "type": "LineString",
    "coordinates": [
      [
        135.03359735012054,
        34.65266643115983
      ]
    ],
    // omit rest of coordinates for
    readability
  ]
},
"category": [],
"refRoad": {"@id": "urn:ngsi-ld:Road
  :675980797"},
"startKilometer": 0,
"endKilometer": 0.064,
"length": 0.064,
"startPoint": {
  "type": "Point",
  "coordinates": [
    135.03360160000003,
    34.652664300000005
  ]
},
"endPoint": {
  "type": "Point",
  "coordinates": [135.0335028,
    34.6521106]
}
}
```

図-2 RoadSegment モデルでのデータ例

た例を図-2、図-3 に示す。Transportation.RoadSegment は道路の地理的な情報（始点、終点等）を含むが、道路の種別（高速道路、一般道等）の情報を持っていないため、Transportation.Road を組みわせる必要がある。各モデルの中でそれぞれ refRoadSegment, refRoad のような形で相互参照しており、道路という情報を2つのモデルクラスで表現している。また、各モデルの location 属性で GeoJSON 互換の情報を保持しているため、GeoJSON としての処理が必要な場合は当該属性を利用すること

```
// Transportation.Road
{
  "@context": [
    "https://smartdatamodels.org/context.
      jsonld",
    "https://uri.etsi.org/ngsi-ld/v1/ngsi-
      ld-core-context.jsonld"
  ],
  "type": "Road",
  "id": "urn:ngsi-ld:Road:675980797",
  "roadClass": "residential",
  "refRoadSegment": {
    "@id": "urn:ngsi-ld:RoadSegment
      :675980797"
  }
  "length": 0.064
}
```

図-3 Road モデルでのデータ例

となる。

なお、道路データを扱う場合はノードとリンクによって構成されたネットワークとして扱うことが多い。一方で、Smart Data Models ではリンクに相当する Transportation.RoadSegment は存在するものの、ノードに相当するモデルが存在しない。これは、現行のモデルクラスが可視化等での利用が中心であり、経路検索等の道路ネットワークが必要なユースケースが多くないためと考えられる。そこで本研究では、ノードに相当するモデルクラスとして鉄道やバスの駅、停留所を表現する Transportation.TransportStation で代用した。

(3) 構築結果および評価観点

図-4 に今回利用したデータ群の可視化結果を示す。地図上に道路や規制情報をマッピングしており、任意の2点を指定すると経路探索が実行される。経路探索の結果は図中の水色の線で表現されており、経路上の高低差や関連する事故情報等を図の下部に表示している。

また、本サンプル実装を用いて以下の観点から提案手法を評価した。

- 評価観点1 ドメイン内の各データを単一のデータモデルで表現可能か
- 評価観点2 ドメイン内のデータに対しドメイン固有の処理が可能か

本サンプル実装ではデータモデル部分にフォーカスしているため、当該部分が関係する機能の評価となる。

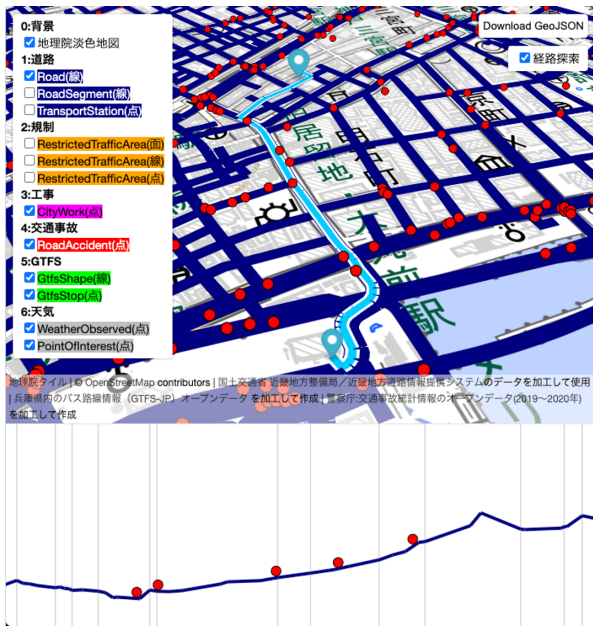


図-4 構築したサービスにおける経路探索の例

表-3 データソースおよび各モデルクラスのデータ属性数 (括弧内の数字は必須の属性数)

データ名	データソース	モデルクラス
道路 (Road)	3 + tags (2)	21 (4)
道路 (RoadSegment)	3 + tags (2)	34 (7)
道路 (TransportStation)	4 + tags (3)	32 (7)
交通規制	224 (-)	23 (2)
工事	15 (-)	61 (3)
事故	58 (-)	38 (3)
バス	17 (5)	13 (-)
天候 (PointOfInterest)	9 (9)	21 (4)
天候 (WeatherObserved)	20 (9)	43 (4)

a) 評価観点 1: ドメイン内の各データを単一のデータモデルで表現可能か

図-4 の経路探索の可視化のように、地図上へ各項目をマッピングすることは可能であったことから、地理空間のマッピングに必要な最低限の情報は表現できていると言える。そこで、以下ではそれ以外の情報の表現について検証する。

表-3 にデータソースと表-2 記載のモデルクラスそれぞれの属性数および必須属性数を示す。表-3 より、交通規制、事故、バス以外は変換先属性数が上回っており、記載可能な情報量が多いことが伺える。一方で、道路に着目した場合、OSM では自由定義可能なタグを任意の数設定できるため、必ずしもモデルクラスがそれらの情報をすべて表現できるとは限らない。例えば、OSM

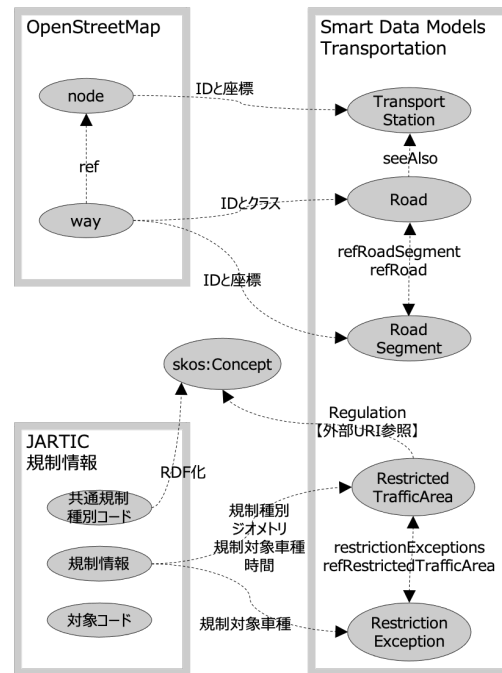


図-5 道路と交通規制を変換した際のデータソース、モデルクラスの関係性

の道路情報には bridge, tunnel といった道路の構造や、立体交差状況を指す layer, 道路の地上, 地下, 水上といったロケーションを指す location といったタグを指定できるが、これらは Road あるいは RoadSegment に対応するものが無いため、変換時に情報が欠落することとなる。

モデルクラスの方が柔軟な例としては、交通規制情報が挙げられる。交通規制情報のデータソースは固定長の CSV で記述されており、タグ等による自由記述の余地は無い。一方、モデルクラス側は規制の内容をリンクで表現する形式となっており、リンク先が情報を追加して様々な情報を表現可能である。

以上のように、緯度経度等の地理情報として最低限必要な情報の表現は可能な一方で、属性の欠落は生じていることが示された。

b) 評価観点 2: ドメイン内のデータに対しドメイン固有の処理が可能か

本評価では、ドメイン固有処理として経路探索を対象とした。事故や規制を考慮しない単純な経路探索自体は、図-4 で示したように道路の GeoJSON から実現可能であった。単純な経路以外の情報、すなわち経路上の規制情報や事故情報は、NGSI-LD, GeoJSON とともに道路へのリンクがされているわけではないため、緯度経度で重畳する必要があった。図-5 に道路情報と交通規制を変換した際の関係性を示す。図-5 より、OSM の道路、JARTIC の交通規制それぞれはモデルクラス間で相互参照できているものの、規制情報と道路が紐付

いていないことがわかる。そのため、利用側において経路探索で得られた経路上の近傍 10m 以内のデータを抽出することでデータの紐付けを実施している。このようなモデルクラス間の参照関係の欠如は可視化や大域的な分析においては特に問題にはならないが、道路が密集している地点において詳細な分析や処理を実行する場合、緯度経度での紐付けは正確性を欠く可能性がある。

直接的な参照関係が無い場合におけるマッチングは、元のデータソースが異なる場合は一般的には解決が困難である。今回のような緯度経度でのマッチングの場合、マップマッチング¹⁹⁾と呼ばれる推定技術が存在するものの、あくまで推定であるため確実な関連付けを保証することは不可能である。モデルクラスへの変換前にマッチングが不要であれば、すなわち今回のようにデータの利用側が推定処理を実施するのであれば自身で手法を選択し、かつ誤差や限界を把握できるため大きな問題とはならないが、モデルクラスへの変換時にマッチングが必要な場合は注意が必要となる。モデルクラス間の参照関係を作るために推定が必要な場合 (e.g., 道路と建物の接続関係等)、推定情報であることの明示や推定精度を併記することが望ましいが、推定に関する情報自体が手法、精度、単位、利用データ等、多岐に渡るため、現状のモデルクラスではすべての情報を保持することは困難である。

以上のことから、ドメイン固有の処理自体は可能である一方で、複数のデータソースのマッチングが課題となることが示された。本課題は提案手法の説明で述べた共通マスタを導入することで一定の解決を得られると考えられるため、同要素を追加した上での検証が必要となる。

5. 考察

(1) 提案手法の適用限界について

以下、本手法およびサンプル実装によって得られた知見の適用可能範囲を述べ、その後評価結果から得られた個別課題への対応について述べる。

まず、本手法で考えるドメインは単一のデータモデルを保持する最小単位としている。既存のデータモデルは対象とする事象は様々あるものの、何かしらの固有処理が必要な単位で作成されているケースが多い。すなわち、GeoJSON では空間に対する演算を IFC では建築物の構造計算を実行できるよう定義されている。多くのドメインで様々なデータモデルが乱立しているが、いずれかの利用の選定と、NGSI-LD のような流通方式の策定自体は可能である。一方で、以下のような点は現状では十分な検討ができていない。

1. データモデルが存在しない事象の扱い
2. データ流通に不向きなデータ形式 (画像, 音声等) の扱い
3. 処理速度や可用性等の運用上の構成

上記の中で、1 および 2 は対象とするユースケースを拡大していくことで順次検証していく必要がある。一方で 3 に関しては、サンプル実装において表示用にタイルベースのデータに再変換したように、ドメイン内、ドメイン間両方で様々な対策が取りうる。また、FIWARE が採用する Orion 等、NGSI-LD を高可用で運用するための OSS 等が存在するため、喫緊の課題とはならないと考えられる。

(2) 個別課題への対応

a) データモデル変換時の情報欠落

データ流通の目的であるデータの利活用だけでなく、法令遵守や特定の情報の保護を考えた場合、情報の欠落自体は許容されるケースがある。例えば、ステークホルダーの機密や個人情報といったデータソースには含まれていても流通は限定的に扱われるデータが存在する。これらのデータは一般的には、完全に秘匿するか、あるいは暗号化された状態²⁰⁾ や匿名加工された状態²¹⁾ で流通させることが望ましい。一方で、対象のドメインで必須とされるような情報が欠落することはドメイン固有の処理の観点から望ましくない。

上記ケースおよび本研究で想定するドメイン単位の処理を考えた場合、ドメイン内で利用されるデータモデルはドメイン内のデータソースのすべての属性を保持する必要はなく、ドメイン固有の処理で必要とされる最低限の属性のみ保持する、という方式が考えられる。当該方式であれば、既存のデータモデル群が固有の処理を可能にする目的もあることから実現は容易である。ただし、単純に情報を欠落させたまま流通を実施した場合、利用者側からは情報がそもそも存在しないのか、あるいは何かしらの事情で欠損あるいは加工されているのかが区別がつかず、データ利活用の障害となりうる。人工知能や機械学習の分野では、データの取り扱いに関しては、Accountability, Transparency²²⁾ が重要だとされており、例えば利活用するデータに対しては入手方法や加工方法を開示する方式が提案されている²³⁾。データ流通が実現した後は機械学習での活用等もみこまれるため、データカタログ¹⁰⁾ や各モデルクラスのリンクとして情報を開示していくことが重要になると考えられる。

b) 異なるデータソースのマッチング

異なるデータソースのマッチングは、様々なケースで課題になりうる。推定によるデータの紐付けの例としては、前述のような緯度経度によるマップマッピン

グだけでなく、監視カメラで撮影された画像から車両の位置や状況を取得したり、人のライフログから行動を推定して他のデータと紐付ける等、様々な状況で生じうる。また、推定が必要な複雑な例でなくとも、製品の統廃合によって製品コードを名寄せする必要性が生じたりする。

上記課題の中で、マスターコードのような問題は共通マスタの管理およびポリシーによる内容の更新、利用の義務化等で対応が可能と考えられる。一方で、推定は手法ごとに様々な結果が存在したり、推定に使ったモデルの差異等もあるため、共通マスタへマスタとなる情報を登録する方式では対応が困難である。前述したデータカタログやリンクを用いてデータがどのように生成されたかを開示することが必要になる。また、推定の方式自体を利用者側が指定または作成することが望ましいケースも考えられるため、データそのものの流通だけでなく、データホルダーと利用者を結びつけるマッチングプラットフォームのような機能²⁴⁾が重要になると考えられる。

6. まとめ

本研究では、複数のステークホルダーがそれぞれ個別に管理しているデータ群を横断的に活用する仕組みを提案した。当該仕組みでは、ドメイン単位で利用するデータモデルを定義し、ドメイン内でのドメイン固有処理の実現とドメイン間のデータ流通を実現する。また、各ドメインで共通して参照される共通マスタを用意することで、ドメイン間での語彙の差異を吸収する。モビリティサービスをユースケースとして有効性を評価した結果、提案した仕組みによりドメイン内での固有処理の利便性を維持しながらデータ流通が可能なが示唆された。一方で、データ項目の欠損や異なるデータソースのマッチングに課題があることが示された。

今後の課題としては、本稿の評価結果で挙げた課題に加えて、ドメイン追加による追加評価および加工済みデータの取扱が挙げられる。今回評価したモビリティだけでなく、オンラインサービスやスマートホーム等、社会には様々なドメインが存在するため、それらのドメインを定義、追加した場合の提案手法が有効かどうかを評価する必要がある。また、本稿の評価では加工前のデータを取り扱っているが、実際のシステムでは規約や個人情報保護の都合で加工済みであったり、統計化された状態のみ流通するデータも存在する。そのようなデータ群をどのようなデータモデルで扱い、流通させるかは規約整備やコンプライアンスの観点で重要な課題となるため、技術面、ルール面両方で検討していく必要がある。

Notes

- ¹ <https://www.openstreetmap.org/>
- ² <https://www.drm.jp/>
- ³ <https://www.jartic.or.jp/>
- ⁴ <https://fed4iot.org/>
- ⁵ <https://imi.go.jp/goi/>
- ⁶ <https://joinup.ec.europa.eu/>
- ⁷ <https://www.w3.org/TR/sparql11-overview/>
- ⁸ <https://www.w3.org/TR/sparql11-federated-query/>
- ⁹ 2022年2月時点で公開されていた定義を利用
- ¹⁰ 例えば CKAN (<https://ckan.org/>) 等。

REFERENCES

- 1) Cirillo, F., Solmaz, G., Berz, E. L., Bauer, M., Cheng, B., and Kovacs, E.: A Standard-based Open Source IoT Platform: FIWARE, *IEEE Internet Things M.*, Vol.2, No.3, pp.12–18, September 2019.
- 2) Abella, A.: Smart Data Models Program: Data models for data interchange, current status and future, January 2022.
- 3) Bauer, M., Kovacs, E., Schulke, A., Ito, N., Criminisi, C., Goix, L.-W., and Valla, M.: The Context API in the OMA Next Generation Service Interface, *2010 14th International Conference on Intelligence in Next Generation Networks*, pp. 1–5, Berlin, Germany, October 2010, IEEE.
- 4) Maas Alliance Working Group 3: Interoperability for Mobility, Data Models, and API, November 2021.
- 5) Open Geospatial Consortium Inc.: OpenGIS Geography Markup Language (GML) Encoding Standard, August 2007.
- 6) Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S., and Schaub, T.: The GeoJSON Format, Request for Comments RFC 7946, Internet Engineering Task Force, August 2016.
- 7) Environmental Systems Research Institute, Inc.: ESRI Shapefile Technical Description, July 1998.
- 8) Open Geospatial Consortium: OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, September 2021.
- 9) Open Geospatial Consortium: OGC IndoorGML: Corrigendum, March 2018.
- 10) ISO: ISO 16739-1:2018 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema, November 2018.
- 11) Hammar, K., Wallin, E. O., Karlberg, P., and Hälleberg, D.: The RealEstateCore Ontology, *The Semantic Web – ISWC 2019*, (Ed. by C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. Cruz, A. Hogan, J. Song, M. LeFrançois, and F. Gandon), pp. 130–145, Cham, 2019, Springer International Publishing.
- 12) Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Berges, M., Culler, D., Gupta, R., Kjærsgaard, M. B., Srivastava, M., and Whitehouse, K.: Brick: Towards a Unified Metadata Schema For Buildings, *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, pp. 41–50, November 2016.
- 13) Google LLC: GTFS Static Overview — Static Transit, <https://developers.google.com/transit/gtfs>, accessed on 2022-03-24.
- 14) NABSA: General Bikeshare Feed Specification, <https://github.com/NABSA/gbfs>, accessed on 2022-03-24.
- 15) Data4PT: Data4PT, <https://data4pt-project.eu/>, accessed on

- 2022-03-24.
- 16) Schema.org Project: Schema.org, <https://schema.org/>, accessed on 2022-03-22.
 - 17) Usuda, Y.: Forefront of Disaster Information Sharing Systems: Supporting Disaster Responses using SIP4D, *IEICE Commun. Mag.*, Vol.15, No.3, pp.192–199, 2021.
 - 18) Detti, A., Nakazato, H., Martínez Navarro, J. A., Tropea, G., Funari, L., Petrucci, L., Sánchez Segado, J. A., and Kanai, K.: VirIoT: A Cloud of Things That Offers IoT Infrastructures as a Service, *Sensors*, Vol.21, No.19, pp.6546, September 2021.
 - 19) Quddus, M. A., Ochieng, W. Y., and Noland, R. B.: Current map-matching algorithms for transport applications: State-of-the art and future research directions, *Transportation Research Part C: Emerging Technologies*, Vol.15, No.5, pp.312–328, October 2007.
 - 20) Evans, D., Kolesnikov, V., and Rosulek, M.: A Pragmatic Introduction to Secure Multi-Party Computation, *FNT in Privacy and Security*, Vol.2, No.2-3, pp.70–246, 2018.
 - 21) Sweeney, L.: K-Anonymity: A Model for Protecting Privacy, *Int. J. Unc. Fuzz. Knowl. Based Syst.*, Vol.10, No.05, pp.557–570, October 2002.
 - 22) Mayernik, M. S.: Open data: Accountability and transparency, *Big Data & Society*, Vol.4, No.2, pp.205395171771885, December 2017.
 - 23) Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., and Crawford, K.: Datasheets for datasets, *Commun. ACM*, Vol.64, No.12, pp.86–92, December 2021.
 - 24) Ramachandran, G. S., Radhakrishnan, R., and Krishnamachari, B.: Towards a Decentralized Data Marketplace for Smart Cities, *2018 IEEE International Smart Cities Conference (ISC2)*, pp. 1–8, Kansas City, MO, USA, September 2018, IEEE.

Multi-Domain Data Interoperability Mechanism Study Based On Mobility Service Usecase

Yukihisa FUJITA, Yuzo MATSUZAWA, Yoichi TAKAHASHI, Masahiro KUWAHARA

With the advance of sensing technologies, various kinds of data are collected and exchanged. When there are various data domains and/or multiple stakeholders, like smart cities, data utilization becomes difficult due to the differences of data format and data granularity. To mitigate such situation, this research proposes a data platform focusing on data interoperability by leveraging OSS and defacto/dejure standards. The proposed method uses data models defined for each data domain and common master data, and they can be accessed through single endpoint. Sample implementation of mobility services usecase implies the proposed method can provide easy data extraction from user perspective.