(12) インフラデータの横断的な連携基盤における検索性向上に関する検討

榎本　真美[1]・石川　雄己[2]・鈴木　仁[2]・関本　義秀[3]

[1]正会員　パシフィックコンサルタンツ株式会社　（〒101-8462 東京都千代田区神田錦町三丁目22番地）
E-mail: mami.enomoto@ss.pacific.co.jp

[2]非会員　パシフィックコンサルタンツ株式会社　（〒101-8462 東京都千代田区神田錦町三丁目22番地）
E-mail: yuuki.ishikawa@tk.pacific.co.jp, hitoshi.suzuki@ss.pacific.co.jp

[3]正会員　東京大学　空間情報科学研究センター（〒277-8568 千葉県柏市柏の葉5-1-5）
Email: sekimoto@iis.u-tokyo.ac.jp

インフラデータの流通ハブとなる「国土交通データプラットフォーム」は，データ連携が拡大する中，様々な機関が作成・提供する大量のデータを利用者が容易に検索，表示，ダウンロードできるよう，データの検索性や一覧性を向上させる仕組みづくりが必要であった．そのため，共通的にメタ情報を機械可読な形式で整備し，検索エンジンの導入，システム連携APIの構築を行うことでそれらを実現した．地方公共団体の工事基本情報を新規にAPI連携するにあたっては，メタ情報を登録して検索可能にするとともに，データ更新の自動化を図った．

Key Words :*Data management, Data platform, API, 3D data, Digital Twin*

1.　はじめに

(1)　背景

内閣府の「第５期科学技術基本計画」では，Society5.0として，「サイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより，経済発展と社会的課題の解決を両立する，人間中心の社会（Society）」を提唱している[1]．この実現に向けて，国土交通省は，フィジカル空間の事象をサイバー空間に再現するデジタルツインにより，業務の効率化や生産性の向上，スマートシティ等の施策の高度化，産学官連携によるイノベーション創出を目指している[2]．

しかしながら，国土交通省では，国土，都市，交通等の多くのデータを保有しており，データは個々の部局ごとに管理・公開されているが．それらを横断的に利活用する基盤がなかった．

そこで，国土交通省では各分野のデータベースやシステムと連携し，分野を超えた横断的なデータ検索や取得を可能とするためのデータ連携基盤である「国土交通データプラットフォーム」を構築することとした．国土交通省自らが保有しているデータと民間等のデータを，分野網羅的および組織横断的に集約し，多種多様なインフラデータの流通ハブの役割が期待されている．

(2)　国土交通データプラットフォームの取り組み状況

国土交通データプラットフォーム[3]は，国土やインフラに関連する各分野のデータを，ウェブブラウザ上の３次元地図で検索・表示・ダウンロードを可能とするサービスとして構築されている．令和２年４月に一般公開を開始し，データ連携数の拡大や機能改良を行い計８回の更新が行われた[4]．



図-1　国土交通データプラットフォーム（Ver2.2）

主たる機能として，利用者は，分野や都道府県を選択してデータの絞り込み，あるいはキーワードによる絞り込みが可能である．絞り込まれたデータファイルが画面の左サイドに一覧表示され，地図上にも示される（図-1）．国土交通省等が保有する道路や河川などのインフラ施設，

工事や地盤，人流に関する多種多様なデータが扱われている．しかしながら，データ連携を拡大する中で，当初計画よりデータ量が増大することが見込まれたため，当初のシステム構成や機能では，ユーザのニーズに十分応えることができず，急激に増大したデータの検索性や，操作性，データ連携の拡張性に問題が生じていた．

(3) 目的

データ連携の拡大に伴い，より高度な検索性を実現するためには，システム構成を見直す必要があった．本稿では，横断的なデータ活用を促進するための「国土交通データプラットフォーム（Ver3.0）」の開発にあたっての技術的検討についてに述べる．多種多様なインフラデータデータを一元的かつ統一的に管理するため，カタログ情報やメタデータの整備を検討し，検索や表示，情報源へのアクセスを容易にすることを目指した．さらに，システム間の連携を実現するためのAPIを導入し，自動的かつ定期的なデータ更新を図った．

2. システムの課題

(1) 検索性の課題と要因

改修以前のシステムでは，ファイルのメタ情報をPostgreSQLデータベースに登録し，表示のフィルタリングやフリーワード検索を行っていた．すべてのデータに統一的に整備されているメタ情報はタイトルのみであった．また，フリーワード検索は，PostgreSQLの全文検索を利用しており，検索機能に制限があるうえに，データ量が増えるにつれてパフォーマンスが低下する要因となっていた．検索結果は，関連度の要素はなくデータベースのインデックス順に表示され，ユーザが期待する検索結果が表示されなかった．

このシステム構成で，検索用メタ情報を統一的に整備した場合，データ量が増えることで処理時間が増大し，パフォーマンスがさらに低下することが考えられた．

(2) システム間連携の課題と要因

約20の機関からデータが提供されていたが，多くのデータ提供者のシステムがＡＰＩなどのデータ連携機能を持ち合わせておらず，データ提供者の都合にあわせた対応を国土交通データプラットフォーム側が行っていた．一部のデータは，ブラウザ上での可視化に不向きであったため，データ変換やファイル形式の変換を行う必要があった．その際に，データセット毎に異なる位置参照系の情報が紐づいておらず，手作業によるデータ確認と変換を余儀なくされた．これらの課題は，API等によるシステム間連携の自動化を進められず，定期的あるいはタイムリーなデータ更を困難にする要因となっていた．

3. 検索性向上の検討

これらの課題を解決するため，以下の検討を行った．

(1) カタログのメタ情報の整備

連携するそれぞれのシステムで作成・管理されるデータベースやファイルは，データ粒度や分類体系が異なる．国土交通データプラットフォームで統一的に扱うため，図-2に示すデータ体系に基づき，カタログとデータセットの情報を整理することとした．

カタログ項目が不揃いであるとデータ管理が煩雑になるだけでなく，利用者が検索しづらくなる．カタログ項目は，W3Cが定義するデータカタログのメタデータにおける標準であるDCAT(Data Catalog Vocabulary)の仕様に基づいて，デジタル庁が別途定義した「DCAT-GoJ」[5]に準拠して整備することとした．

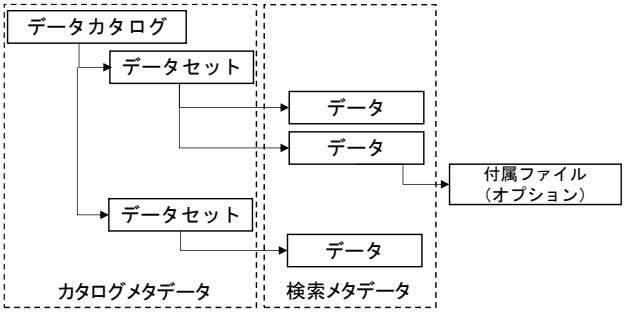


図-2 データ体系図

表-1 カタログメタデータの抜粋

項目	項目名	データ形式	必須
Identifier	管理 ID	文字列	○
Title	タイトル	文字列	○
Dataset	データセット	自由文字列	○
Description	説明	自由文字列	○
Theme	テーマ分類	改行リスト	○
Tags	タグ	改行リスト	○
Spatial	対象地域	文字列	
publisher.organization	データ提供者の組織名	自由文字列	○
publisher.homepage	データ提供者の HP	URL	
Issued	公開日	日付	
Modified	最終更新日	日付	
license	ライセンス	改行リスト	○
Homepage	ホームページ	URL	

カタログやデータセットのメタデータを共通的に整備することにより，利用者にデータの提供者情報や，データ更新日，ライセンスなどの情報を提供が可能となった．

(2) データの検索メタデータの整備

キーワード検索の結果を迅速に表示するため，各データの検索メタデータを適切に管理しつつ，効率的に検索・抽出ができる仕組みが必要であった．加えて，データ提供機関によって異なるデータ特性やデータの持ち方に応じて，運用負担が少なく，より効率的かつ持続可能性の高いメタデータ整備が必要であった．共通のメタデータとして最低限の情報を整備することで、データ管理に必要な情報と検索・処理に必要な情報のデータ内容について統一を図った．

まず，データ管理に必要な情報として，各データのユニークID，作成日時，最新更新日時，削除日時が要求事項として挙げられた．データの作成日や更新日など，関連した情報が適切に保持されることで，利用者がデータを発見・利用できるようになる．これらの管理データが存在しない場合は，連携するシステムの更新を行う際に，全件データを再取得および再登録する必要があり，システムへの負担が大きくなる．

さらに，データ内容に関するメタデータとして，データの名称やカテゴリー，地理情報，日付情報を統一的にデータ収集することとした．これにより，異なるデータセットとの横断的な検索が容易となる．また，連携システムに応じて，データ項目やデータ型にバラツキがあるため，スキーマに対応付けて変換（マッピング）を行い，相互運用性，一貫性，均質性が保たれるようにした．

(3) 検索エンジンの導入

効率的かつ迅速な検索と検索結果の表示が求められる．そこで，高性能で柔軟な検索エンジンOpenSearchを導入することとした．前項で整理した共通メタデータと，データセット毎の独自項目を検索対象フィールドとした．

国土交通データプラットフォームのアプリケーションGUI上で，検索項目を指定して検索フィルターを作成し，複数項目の検索条件を組み合わせた高度な検索が可能となった．また，各データの地理情報を用いて，任意の緯度経度からの距離で検索を可能とした．これにより，ユーザが画面表示する地図範囲内においてデータを抽出することが可能となる．

また，類義語を定義することにより，ユーザが意図した語句が完全に一致しなくても検索結果がヒットする機能を追加した．例えば，「羽田空港」と入力した場合にも「東京国際空港」という名称が含まれるデータも検索結果に表示される．施設名称の略称や愛称など，約350件の類義語の辞書を作成して登録した．加えて，サイトへのアクセス解析結果を用いて，頻繁にフリーワードとして検索されている単語を選定し，それを考慮に入れることで検索機能を改良した．例えば「プラトー」と入力して検索を行った場合に，「PLATEAU」をデータセット名に含むデータが検索結果として出力される．

(4) システム連携API

データ提供元サーバから国土交通データプラットフォームのサーバへデータを送信する際にはREST-APIを使用する方法とファイル自体でやり取りをするSFTPの２種類の接続方式を用意した．

データ提供者との連携を効率良く統一したプロセスで連携するための汎用的なAPIを提供することとした．データ連携は基本的にはAPIを使用する方法で連携を行うものとするが，REST-APIは大容量データの送信には適さないため，大容量データの送信を行う場合や連携先でREST-APIを使用することができない場合は，SFTPを用いてデータのやり取りを実現した．ファイル容量が大きい場合や，リアルタイムなファイル同期が必要な場合には，SFTPによる連携が推奨される．

4. 地方公共団体の工事データの連携

地方公共団体による工事基本情報を対象として，前章までに述べた方法で，実際の新規システム連携を行った．地方公共団体向けのオンライン電子納品システムのサービスの１つであるMyCityConstructionに登録された工事の

表-2 地方公共団体の工事基本情報のデータマッピング 抜粋

MyCityConstructionのデータ		国土交通データプラットフォームへ登録するメタデータ					
属性	データ型	フィールド名（オリジナル）	DPF:フィールド名	表示名	データ型	検索対象項目	キーワード検索対象項目
工事名称	str	MCC:construction_name	-	工事名称	String	○	○
		-	DPF:title	名称	String	○	
発注機関名	str	MCC:client_organization_name	-	発注機関名	String	○	
受注機関名	str	MCC:contractor_organization_name	-	受注機関名	String	○	
工事写真	str	MCC:construction_photo	-	工事写真	String		
工期開始日	str	MCC:construction_start_date	-	工期開始日	String	○	
工期終了日	str	MCC:construction_end_date	-	工期終了日	String	○	
		-	DPF:year	年度	String		
		-	DPF:completion_datetime	完成日	Timestamp	○	
工事内容	str	MCC:construction_content	-	工事内容	String	○	
coordinates	array	-	DPF:longitude	経度	Float	○	-
		-	DPF:latitude	緯度	Float	○	-

基本情報を，API連携により国土交通データプラットフォームと連携させた．工事受注者によりオンライン電子納品システムにアップロードされた工事成果品を工事発注者が確認及び承認すると，工事名や場所などの基本情報はオープンデータとしてウェブ上に一般公開される．そのタイミングで，国土交通データプラットフォームのAPIを呼び出し，工事基本情報のデータ登録や更新が実行される．表-2に示すデータマッピングに応じて，国土交通データプラットフォームに検索メタ情報の登録を行った．国土交通データプラットフォームの利用者は，ブラウザ上で，工事基本情報の検索，表示，ダウンロードが可能となった．（図-3）



図-3 地方公共団体の工事基本情報の表示例

5. 考察

国土交通データプラットフォームは，データ流通のハブとなるプラットフォームであるため，ユーザの幅が広く，様々な目的やデータ熟度の利用がある．インフラデータにおいては，計画段階から設計，工事，供用，維持管理といった何十年にも渡るライフサイクルの中で，データを利活用することも求められる．これらの目的に対応するためには，様々な機関によって作成され蓄積される大量の情報に，横串を通して，容易に利用できる仕組みを確立することが必要であった．そのため本検討では，国土交通データプラットフォームで取り扱うデータを，共通的にメタ情報を機械可読な形式で整備した．その結果，ユーザが求める様々な検索パターンに対応し，柔軟かつ的中率の高い検索機能を提供することができた．今後も，利用状況のログの適切な分析を行い，よく検索されるキーワードや条件など利用者ニーズを捉えたチューニングや類義語辞書の充実化を行うことで，性能向上が期待できる．

前述した機能面の改良と合わせて，品質管理を適切に行い運営していく必要があるが，同時に，データ利用者がデータの特性や品質を理解して活用できるよう，データの鮮度や精度を分かりやすく伝えていくことが重要である．そのためには，データ提供者と協調して，適切なシステム間連携を継続し，適切な頻度とタイミングで更新を行えるよう，連携処理の自動化を進めることが必要である．

6. まとめ

データのオープン化，データ基盤の整備やシステムのリプレイスが進む中，国土交通データプラットフォームはインフラデータのハブとしての役割がますます期待されている．

今後の課題として，システム間連携を拡張しながらも連携を円滑に進めるために，標準的かつ汎用的なデータ連携仕様として，データ要件と連携要件を定義する必要がある．また，利用者がブラウザ上で操作し検索やデータダウンロードを行うGUIの機能だけでなく，APIの提供も重要である．民間等によるAPIを用いた様々なソリューションやアプリケーションが創出され，データ利活用が促進されるよう，データ連携拡充やデータ流通基盤としての機能向上を目指す．

謝辞：本研究成果は，国土交通省国土技術政策総合研究所からの発注業務により得られた成果の一部であり，国土技術政策総合研究所及び国土交通データプラットフォーム関係者の方々に多大なるご協力を頂いた．ここに謝意を表する．

参考文献

1) 内閣府：Society 5.0, <https://www8.cao.go.jp/cstp/society5_0/>,（入手 2023.6.1）.
2) 国土交通省：国土交通データプラットフォーム（仮称）整備計画, <https://www.mlit.go.jp/common/001291151.pdf>,（入手 2023.6.1）.
3) 国土交通省：国土交通データプラットフォーム, <https://www.mlit-data.jp/>,（入手 2023.6.1）.
4) 伊藤太一，中村英佑：国土交通データプラットフォームの構築に向けた取組～“すべてのデータを一つに”を目指して～，土木技術資料, Vo.65，No.1, pp.10-13, 2023.
5) デジタル庁：メタデータ導入実践ガイドブック,（入手 2023.6.1）.