

汎用的な人流データのフォーマット(案)の検討について

令和5年3月
国土交通省 不動産・建設経済局

汎用的な人流データのフォーマット(案)の検討について

- 目的

- 人流データの取得手法の多様化・高度化が進み、人流の把握や分析にもとづく将来予測等が容易となっている一方で、人流データの提供者や計測手法によりデータ形式やファイル形式等が異なり、データ仕様や特徴を理解することが難しく、そのことが人流データの流通を阻害している。また、取得した人流データをうまく可視化し表現することができず活用まで至らないケースも多々存在している。このため、現状流通しているデータについて調査し、汎用性のあるデータ形式等について検討案を作成することとした。
- 本検討結果は事業者による人流データの自由な提供や競争を妨げるものではなく、この検討結果を参考にして事業者あるいは業界団体等による検討を促進することができればと考えている。
- また、本検討結果にもとづき、人流データ可視化ツールの試作開発を行う。

- 検討方法

- 流通している人流データや公開されている人流計測データに関する情報を基に、包括的にこれらを扱うためにどのようなデータの持ち方をすべきかの検討を行った。
- 人流データの想定利用者である自治体職員や人流データベンダー等へのヒアリングを行い、双方の意見を聞きながらフォーマット案を検討した。

汎用的な人流データのフォーマット(案)の検討について

• 検討に際して考慮した点

- 人流データにおいては計測された人数そのものと同等に、それがいつ、どこで、どのようにして、どんな人を対象に取得されたのかという付帯情報(メタ情報)が重要であると考えた。このため大きく、「人流データ」「人流データを補完するデータ」の2つに分けて整理した。
- 人流データの中で最も粒度が細かいものは、個人の位置を連続的に記録した移動軌跡データであり、GPSで取得されるデータなどがある。このようなデータから必要な処理を施した上で生成されるのが、ある地点やエリアにおいて集計されたデータであり、これをポイント計測データと、幾何図形集計データとして分類した。
- ポイント計測データは移動軌跡データからの集計以外に、設置されたセンサーによる計測値が該当する。また、幾何図形集計データの代表的なものがメッシュ統計データである。
- さらに、人流データの利用目的としてニーズが高いのが人々の回遊調査であり、この基礎的な情報としてOD表の基となるODデータのフォーマットを検討した。

• 課題

- ここでは人流のデータ形式における項目や並び順などの提案にとどめ、その中身のデータの定義やデータの品質(精度や統計的な信頼度)までは規定することはせず、まずは各社が自由に定義できる人流データを汎用的に扱えるフォーマットを目指した。より利用者が扱いやすく、各種データを比較検討できるようにするためには、単位や属性項目、精度などが揃っているとよいが、人流データのソースが異なることや各社の創意工夫によるアルゴリズムの差などがあり、これらを統一することは今後さらなる検討を要する。
- ODデータについては交通調査や観光分析など、目的に応じて集計時間幅等が異なることや、ユニーク化(一定時間やエリアにおいて同一人物をどのように集計するか)の考え方、滞在時間の算出方法など実際にデータを作成する際には多くの検討事項がある。

汎用的な人流データのフォーマット（案）

種別	データ名称	データの説明	列名	列型	列の説明	サンプル
人流データ	①移動軌跡データ	GPS等で得られたヒト、モノ(ID)等の移動軌跡	id	文字列型	一意のユーザーを特定するID	12345
			timestamp	datetime型	データ計測時間	2020-06-02 23:57:12.120174 JST
			lat	実数型	移動体の位置情報(緯度)	35.70598
			lon	実数型	移動体の位置情報(経度)	139.66576
			height	実数型	移動体の位置情報(高さ 単位: m)	10.5
			value	実数型	補正や拡大推計などの係数に使用	20.5
			option1~	文字列型	任意で設定する付帯情報(市区町村コード、移動速度、曜日など)	13001
	②ポイント計測データ	設置されたセンサー等のポイントに対して、得られた計測値	id	文字列型	一意のポイントを特定するID	12345
			timestamp	datetime型	データ計測時間	2020-06-02 23:57:12.120174 JST
			value	実数型	計測値(実測値若しくは拡大推計値)	10
			option1~	文字列型	任意で設定する付帯情報(移動手段、方向など)	自転車
	③幾何図形集計データ	①から特定エリア内で集計されたデータ	id	文字列型	メッシュ番号等の識別子	5037734241
			value	実数型	計測値(実測値若しくは拡大推計値)	1000
			option1~	文字列型	任意で設定する付帯情報(性別、年代、都道府県名、居住区分など)	男性
	④ODデータ	①から生成された一定時間内の起点終点データ	time_o	datetime型	検知した日時(起点)	2020-06-02 23:00:00.00000 JST
			time_d	datetime型	検知した日時(終点)	2020-06-02 23:15:00.00000 JST
			origin	文字列型	起点の名称またはエリアID	ABCDE
			destination	文字列型	終点の名称またはエリアID	VWXYZ
			id	文字列型	一意のユーザーを特定するID	12345
			value	実数型	計測値(実測値若しくは拡大推計値)	52.18
			option1~	文字列型	任意で設定する付帯情報(移動手段など)	女性

種別	データ名称	データの説明	列名	列型	列の説明	サンプル
人流データを補完するデータ	⑤ユーザー属性データ <人の属性>	①④のidについての属性情報	id	文字列型	一意のユーザーを特定するID	12345
			option1~	文字列型	ユーザーについての付帯情報(性別、年代など)	男性
	⑥地点情報データ <場所の属性>	②のセンサー等の位置情報、④のorigin/destinationにあたる地点の情報	id	文字列型	一意のポイントやエリアを特定するID	ABCDE
			lat	実数型	緯度	35.70598
			lon	実数型	経度	139.66576
			height	実数型	高さ(単位: m)	20.0
	option1~	文字列型	idに対して任意に設定する付帯情報(地点名称など)	東京駅		
	⑦ジオメトリデータ	③のポリゴンデータなどの位置情報等	id	文字列型	メッシュ番号等の識別子	5037734241
			geometry	文字列型	ジオメトリ情報	{ "type": "Polygon", "coordinates": [[[139.0625, 35.23333333], [139.075, 35.23333333], [139.075, 35.24166667], [139.0625, 35.24166667], [139.0625, 35.23333333]]] }
			option1~	文字列型	任意で設定するポリゴンデータごとの付帯情報	area1
	⑧オプションメタデータ	①~⑧にもつ各種オプション項目(option1~)の意味を説明する名称	option	文字列型	対応するオプション番号	option1
name			文字列型	オプションが意図する内容(option1~列の別名)	市区町村コード	

汎用的な人流データのフォーマット(案)の検討について

- 作成フォーマット(案)の説明
 - 計測値(value)を持つ「人流データ」と、「人流データを補完するデータ」の2種に大別した。
 - 「人流データ」のフォーマットとして以下の4種を用意した。
 - ①移動軌跡データ
 - ②ポイント計測データ
 - ③幾何図形集計データ
 - ④ODデータ
 - 「人流データを補完する情報」は、計測対象<人の属性>や計測地点<場所の属性>、ジオメトリ情報を付与するために用いるもので、各idをキーにして2つのデータを紐づける。
 - 各フォーマットにおいては1つの数値情報(value)と、最低限の項目に絞って定義し、その他の付帯情報はoptionという形で自由度をもって利用できるようにした。
- option項目について
 - 先に述べたように人流データベンダー各社から提供されるデータに持つ項目(曜日、属性、計測対象、スマホのOS等)はさまざまであり、これらを汎用的に扱うためにoption項目とした。
 - optionで任意に設定する内容は、緯度経度に対する市区町村コードのような情報であったり、独自のアルゴリズムから生成された移動速度、あるいは日付から得られる曜日など、人流データを分析する際にあらかじめ保持しておく役に立つ情報などとして自由に利用できる。
 - これらのoptionの持つ意味を「人流データを補完するデータ」(下の表)にある⑧オプションメタデータで定義する。