

センシングデータ流通市場におけるメタデータの定義・生成・活用の一方式

An Approach of Defining, Generating and Utilizing Metadata for Sensing Data Trading Market

小田利彦^{*1}
Toshihiko Oda

今井紘^{*1}
Hiroshi Imai

内藤丈嗣^{*1}
Takeshi Naito

竹林一^{*1}
Hajime Takebayashi

^{*1} オムロン(株) 技術・知財本部 SDTM 推進室

OMRON Corporation, Center for Sensing Data Trading Market, Technology and Intellectual Property H.Q.

Demand for sensing data trading market (SDTM) is increasing as becoming more important of application developers who use sensing data gathered from various IoT devices. For them, metadata of sensing data is crucial information to analyze it especially with AI technology. This paper discusses metadata definition, generation and utilizing technology with implementation on prototype.

1. はじめに

さまざまな場所に存在する IoT デバイスがネットワークに接続され、そこで生成されたセンシングデータを中心とするデータがクラウド上にビッグデータとして収集・蓄積されつつある。こうしたセンシングデータの多様な組み合わせによって、画期的なサービスが数多く生まれる期待がある。そのためには、センシングデータを使ったサービスを開発・提供するデータ利用者が、安全かつ容易にデータを入手できる流通市場の重要性が高まっている(図 1 参照)。

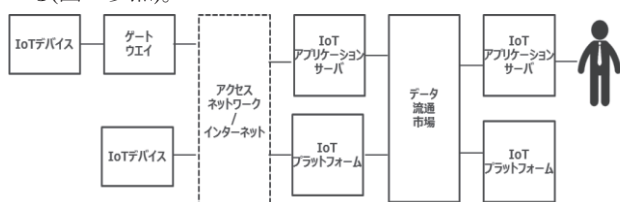


図 1 IoT システムとデータ流通市場

著者等は、センシングデータデータ流通市場(SDTM: Sensing Data Trading Market)の確立を推進しているが、SDTM を介してデータ利用者がセンシングデータを検索・マッチング・活用するためには、センシングデータだけでなくそのメタデータも流通させることが重要であると考え。本論文では、データ流通市場の展開を背景にして、メタデータの形式表現の定義、メタデータの生成・配信、メタデータを活用する方式について、課題とその対策、そしてプロトタイプによる評価を述べる。

2. センシングデータのメタデータ

メタデータとは、「データに関するデータ」のことである。センシングデータを分析し処理するために有用とされるメタデータの種類を以下に示す。

- センサの識別情報:
製品名、モデル型式、ロット番号、FW バージョン等
- センサの性能仕様に関する情報:
計測精度、分解能、計測限界範囲、時刻精度 等
- センサから出力されるデータ仕様:
データ単位系、小数表現、区切り文字 等
- センサの設置条件に関する情報:

- センサの設置位置、高さ、治具、設置者、補正有無 等
- 観測対象情報に関する情報:
センサが観測する対象についての属性情報で、
血圧計の例では、測定者の年齢、性別、病歴、運動量等

3. メタデータの定義

センシングデータのメタデータは、SDTM を介してリアルタイムに配信される場合がある。そのため人が読むのではなく、コンピュータで処理できる形式とする必要がある。さらに、データ利用者が複数のデータ提供者からデータを入手しても統一的にデータ解釈するためには、入手先が異なっても共通の構文や語彙を用いて記述されていることが課題となる。

そこでセンシングデータのメタデータを記述する構文や語彙として、国際的な技術標準化団体である Open Geospatial Consortium の技術規格(以降 OGC 規格と呼ぶ)に基づくこととした。この OGC 規格には、センサによる計測に関する情報や機器の仕様に関するデータモデル及び XML 実装が公開されている[OGC 2011,2012,2014]。

また、2 章で示したメタデータの種類を、OGC 規格のデータモデルで表現したが、そこに含まれない観測対象に関する情報やセンサの設置条件情報などは、OGC 規格で定義された Physical Component クラスをサブクラス化して追加することにした(図 2 参照)。

メタデータの語彙は、できる限り公開された共通語彙を参照することが望ましく、単位系の公開共通語彙や、DBpedia に登録されている語彙を用いた。一方、特殊な語彙として、センサ製品に固有の出力データ項目や性能項目などの語彙に対して、その意味、表記名、ユニークコードとなる URL を語彙辞書に登録した。この語彙辞書を公開することで、データ利用者はセンシングデータのメタデータにアクセスできる。

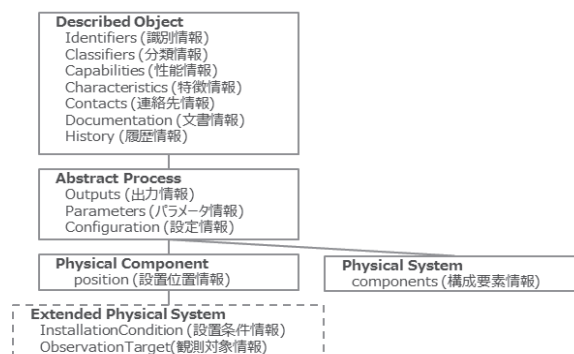


図 2 OGC 規格に基づくメタデータのデータモデル

連絡先:小田利彦、オムロン株式会社 技術・知財本部 SDTM 推進室、〒619-0283 京都府木津川市木津川台 9 丁目 1 番、
toshihiko_oda@oss-g.omron.co.jp

他方、センシングデータのメタデータの表現として、概念の意味ネットワークを構成する RDF を用いた研究が多くある [Qanbari2015]。今回は、実用性を考えることで、クラウド上でスケールできる RDF リポジトリサービスが無かったことや、SPARQL でアクセスする技術普及が乏しいことから、メタデータ表現は XML 形式を評価することとした。

4. メタデータの生成・配信

2 章で示したメタデータは、センサの製品情報のような変化しない静的メタデータと、設置環境やデバイスの状態等の時系列変化する動的メタデータとに分けられる。

静的メタデータは、センサメーカーやデータ提供者が作成してダウンロードサイトに登録しておくことで、利用者がアクセスして入手できる。一方、動的メタデータは、センサやゲートウェイ (GW)、IoT サーバ等において生成され、データ利用者に送信される (図 3 参照)。

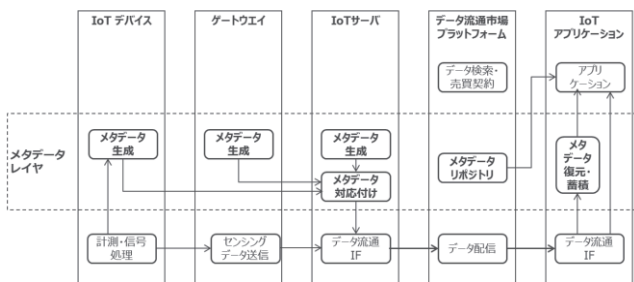


図3 メタデータの生成・配信の流れ

しかし、計算リソースが少ないセンサや GW から、OGC 規格の冗長性がある XML 形式のメタデータを生成配信することには、リソース上の課題がある。

そこで、メタデータを XML データではなく圧縮した形式で生成・配信を行い、データ利用者に渡す直前で復元する方式とした。

また、データ提供者が定義したメタデータを、自身がプログラムを作成しなくてもメタデータ生成が実行できる仕組みが課題となる。

そこで、データ提供者がメタデータをエクセル上で記述することで、その情報に基づきメタデータ生成モジュールの入力となる変換定義やテンプレートが自動生成される。これによりノンプログラミングで、メタデータ生成がカスタマイズされて実行できる。

5. メタデータの活用

センシングデータを利用したアプリケーションサービスの開発では、センシングデータの AI 分析や統計解析が行われる。その際、データ分析者は、第 3 者から入手したデータをデータクレンジングする前処理作業に労力がかかっており、その軽減が

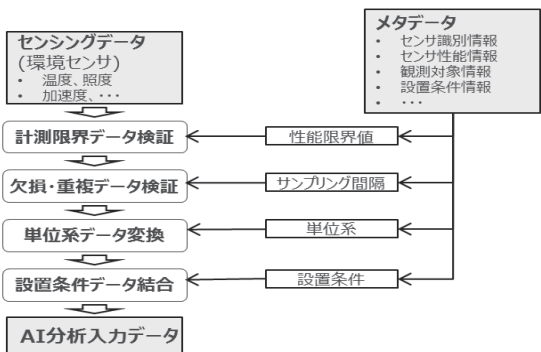


図4 メタデータ活用の前処理の例

課題となる。

そこで、メタデータの活用の 1 つとして、この前処理の作業を支援するツールによって自動化することが有効と考え、次の 3 種の前処理を取り上げて、ツール化を行う。

- ① データ検証: データの異常値や抜け・重複を、除去や補間
- ② データ変換: データの単位系、日時表現、カテゴリの統一化
- ③ データ結合: 異なる入手先から得たデータ同士を結合

6. 実験評価

評価実験として、複数の住居内に環境センサを設置して、計測されたセンシングデータを収集する IoT システムを構築した。このシステム上に、メタデータの生成・配信・活用までのプロトタイプを実装して評価を行った。

メタデータの定義では、本実験において洗い出されたメタデータを、OGC 規格を拡張した構文と語彙辞書に基づき、全てを記述できた。

メタデータ対象	メタデータ項目	内容	タイプ
IoTサーバ	識別情報	正式名称	静的
	識別情報	正式名称、略称、製造者、モデル型式等	
	識別情報	機器識別番号	
	位置情報	経度、緯度	
ゲートウェイ	識別情報	正式名称、略称、製造者、モデル番号等	静的
	分類情報	センサータイプ、利用用途	
	接続先情報	組織名称、接続先等	
	文書情報	マニュアル、製品仕様書	
	性能情報	精度、分解能、計測限界、計測間隔等	動的
	出力情報	出力値の名称、単位、定義小数桁数等	
	パラメータ情報	イベント閾値、パラメータ名等	
	識別情報	機器識別番号、FWバージョン	
	パラメータ設定情報	最大閾値、最小閾値、平均最大閾値等	
	設置条件情報	住居タイプ、世帯人数、設置場所等	
	観測対象情報	性別、年齢、使用機器種類等	
	データ欠損情報	開始番号、終了番号等	
	センサ時刻設定情報	センサ時刻設定日時情報	
	センサ状態情報	温度センサ異常、CPU異常等	

図5 IoT実験におけるメタデータ一覧

メタデータ生成・配信では、センサ自体でメタデータ生成は困難であったため GW で肩代わりする対応を行い、さらにメタデータの XML 記述を断片化して送信することで通信量を許容レベルまで縮小できた。

メタデータの活用では、IoT システムで収集した環境センサのデータから人の行動判別する AI 学習の前処理において、メタデータと連携してデータ検証、データ変換、データ結合を行う処理を作成し、Jupyter notebook 上で評価した。

評価結果として、データ分析者は、前処理する際に、メタデータを使うことは必須ではあるが、入手できれば簡単なスクリプトを作成して対応できる。従ってデータ分析者には、メタデータを取得する IF を整備するニーズの方が高いといえる。しかし、IT スキルが不足する分析者には、逆に、前処理をノンプログラミングで行えるレベルまでツール化することが必要となると判った。

7. 今後の課題

データ流通において、メタデータも重要な流通対象と考えて定義や生成・配信、活用の課題と解決策を示した。

メタデータの定義に関しては、民間主導の協議会等においてメタデータ記述方式の標準化を議論していく予定である。

また、複数のセンシング分野に適用範囲を広げ、メタデータ活用の改善効果を高めるトライアルしていきたい。

参考文献

[OGC2014] SensorML: Model and XML Encoding Standard
https://portal.opengeospatial.org/files/?artifact_id=55939

[OGC2012] Sensor Observation Service Interface Standard:
http://portal.opengeospatial.org/files/?artifact_id=38476

[Qanbari2015] S. Qanbari et al: Linked Sensed Data Enrichment and Analytics Middleware for IoT Gateways, FiCloud 2015, August 2015