

記述形式の自動変換に基づく異種データ連携における
型定義の自動化に関する研究
大谷英之

**A Study on Automated Type Definition in the System for Utilization of
Heterogeneous Datasets Based on Automatic Conversions of Data formats**
Hideyuki O-TANI

Abstract: Recent data explosion and rapid enhancement of computer power have been raising a question about the feasibility of data format standardization to share and utilize data and programs effectively. Once a standard format is defined, many programs depends on it, and the cost of replacing the standard format with a new one becomes extremely high. A solution to this problem is enhancing program reusability by implementing automatic format conversions from the new format to the old format with automatic application of programs which input the old format. In this study, we have developed a method to incorporate data type definitions from XML to the system for automatic conversions of data formats. This development enables us to utilize the numerous heterogeneous data formats and programs effectively.

Keywords: 異種データ (heterogeneous datasets), 自動型定義 (automatic type definition), XML スキーマ (XML schema)

1. はじめに

複合災害とその影響を想定する総合的な防災・減災の実現のため、都市の詳細な描像に基づいて地震・津波等の災害、経済への影響などを検討する研究が進められている (Muneo Hori (2018)) が、これには幅の広い学際的な協力が必要である。一方で、都市のデータは増え続け、計算技術の進展も早い。異種データ・異種プログラムを柔軟に追加・取り替え可能な疎結合の状態連携・共有する方法論が不可欠である。

この方法論の一つが「記述形式の標準化」であり、すべてのデータを標準化された汎用的な形式

で記述することで、標準に対応したプログラム間のデータ連携を可能とする。しかし、この方法論は、計算効率の最適化に特化したデータ形式の採用が必須となる大規模数値解析プログラム間でのデータ授受には適切でない。

そこで、著者らは、一律的な形式でデータを記述する標準化に替わり、記述形式の自動変換に基づくデータの抽象化というより包括的な方法論に基づいて、異種データ・異種プログラム群を柔軟に連携させるデータ処理プラットフォーム (DPP) というインタプリタおよびインタプリタ言語を開発している (図-1)。

本研究は、DPP による記述形式の自動変換が要求する DPP へのデータ型 (記述形式) の登録を、すでに XML スキーマによって記述された定義を読

大谷英之

理化学研究所 計算科学研究センター

E-mail : h. o-tani@riken. jp

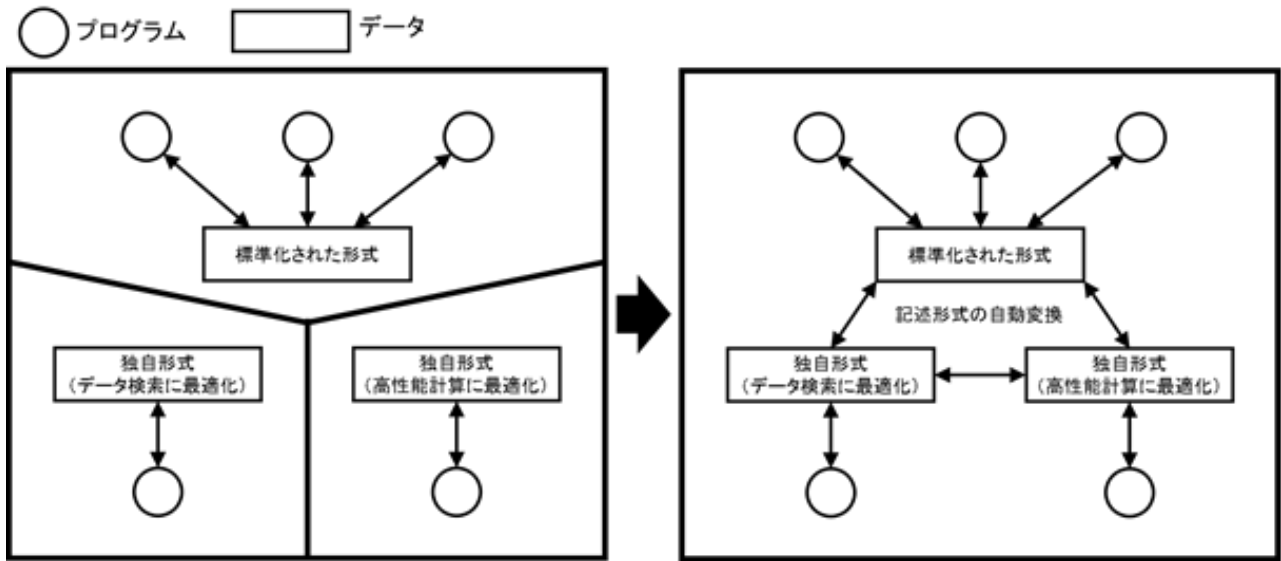


図-1 記述形式の自動変換に基づくデータ・プログラムの孤立化の解消

み込むことによって自動化し、手動で型定義をする場合とはオーダーの異なる数のデータ型を DPP で取り扱い可能とすることで標準化されたデータ形式と個別開発されるデータ形式との連携を強化することを目的とする。

このため、本研究では、まず DPP に対して、JPGIS (GML) 形式のデータ型を XML スキーマによる記述に基づいて登録する。そして、DPP に登録した JPGIS (GML) 形式のデータ型間の変換を XML スキーマ言語で記述された継承関係に従って自動定義し、最後に、JPGIS (GML) 形式と種々の個別開発されたプログラムがもつ独自形式のデータとの間の変換を定義する。

2. 節では、DPP の詳細について説明し、3. 節で DPP のデータ型を XML スキーマの記述に基づいて自動定義する手法の概略を説明する。そして、4. 節で DPP 型を自動定義する試行とその結果について記し、5. で本研究をまとめる。

2. データ処理プラットフォーム (DPP)

2.1 DPP によるデータの自動変換

DPP は、データの記述形式が機械部品の規格などと異なり、本来容易に変換できることを応用し、

オブジェクト指向プログラミングの is-a 関係にあたる関係の成立を継承関係の条件とし、オブジェクト指向プログラミングのアップキャストにあたる変換を自動で行う。このとき、変換開始時のデータ型から変換後のデータ型へ複数のデータ型を辿りながら変換される経路は変換のコストが最小となるように自動選択される。また、特定の条件が満たされた場合にダウンキャストを自動実行することも可能としている。現在、著者らの知る限り、こうした自動経路選択と自動変換が可能なプログラミング言語は DPP のみである。

2.2 DPP データ型の具体的実装

インタプリタである DPP のスクリプトで利用できる関数は、C++言語で作成した関数をラップして作成でき、DPP の制御部は多数の C++言語の関数へのポインタを管理・運用するが、新しい DPP データ型の導入ごとに制御部をコンパイルし直さなくて済むように、DPP が管理する C++言語の関数へのポインタの型は統一している。ここで C++言語の関数の引数および戻り値のための統一的な型として、NS_DC::Object を定義している。この型は、C++言語の任意のクラスのインスタンスが格納できるよう設計しているため、引数を統

一的な型に限定しても関数を表現する能力は落ちない。

3. DPP 型の自動定義方法

3.1 XML のための DPP データ型

DPP のスクリプトで利用できるデータ型は、すべて前述の NS_DC::Object に格納する C++言語のクラスを定めることで定義されている。XML に含まれるデータを表す DPP のオブジェクトは、XML 文書の断片を保持できる C++関数のクラスを定義し、NS_DC::Object に格納することで作成する。

3.2 DPP データ型の自動定義

XML の字句解析と構文解析を行うツールとして Boost ソフトウェアライセンスで提供されている RapidXML を利用する。DPP は、XML スキーマファイルの xs:simpleType 要素 xs:complexType 要素の記述を抽出し、データ型の情報として保持する。ここで、記述の抽出は、前述のように単にそのまま XML 文書の断片として保持すれば、情報の不足が生じないため十分である。ただし、XML の名前空間 URI は正しく記録しておく必要がある。

3.3 XML スキーマに記述されたデータ型の継承関係に基づく DPP データ型の継承関係および自動変換の設定

XML スキーマで定義されたデータ型間には xs:restriction 要素や xs:extension 要素を利用して継承関係が定義される。この継承関係の定義から DPP の継承関係と変換関数を自動定義することが可能である。変換関数は一般的に自動定義が難しいが、XML に含まれるすべてのデータは C++言語の同じクラスを格納した NS_DC::Object で統一的に表現するとしているため、恒等変換、つまり変換しないことを変換関数の中身とすることができる。XML スキーマで記述された継承関係を DPP に対して自動定義可能となることによって、国際標準の XML データのデータ型に対して定義された処理関数が国際標準のデータ型を継承した国内標準のデータ型に対しても自動変換を通し

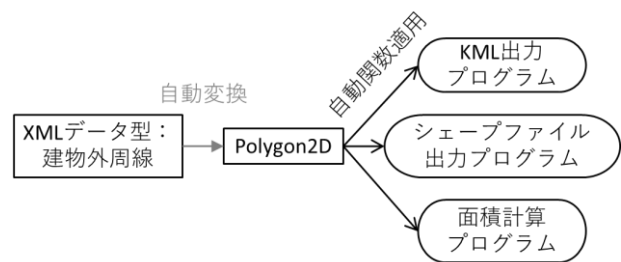


図-2 XML データに対する自動データ変換と自動関数適用の例

てただちに利用可能となる。より、具体的には、GML 形式で記述された地物に対して可視化の処理プログラムを作成しておけば、国内標準の JPGIS (GML) 形式の XML スキーマと XML データのセットに対してただちに可視化が可能となる。

3.4 XML のデータ型と XML を利用しないデータ形式との間の変換の定義

DPP において、XML のデータ型と XML のデータ型間の変換関数は XML スキーマファイルを読み込むことで自動定義が可能となるが、その他のデータ型については、データ型の詳細を把握して手動でプログラムを作成する必要がある。ただし、いったん作成したプログラムは DPP によって効率的な共有が可能であり、国際標準のデータ型と国内標準のデータ型に正しい継承関係が設定されていれば、国際標準に対して定義された便利な処理関数を各国の標準に対してもただちに利用可能となる。

4. DPP 型の自動定義の試行とその結果

基盤地図情報で提供されている地理情報の国内標準である JPGIS (GML) 形式の XML スキーマファイルと、そのスキーマファイルに従って記述されている建物の外周線の情報に対して本研究の手法を適用した。

JPGIS (GML) 形式の XML スキーマファイルからデータ型の情報を抽出してデータ型を自動定義する部分については、すべてのデータ型と継承関係を抽出することに成功した。図-2 に示す通り、

DPP には、多角形状の地理情報のための DPP データ型 Polygon2D をあらかじめ定義し、そのデータを Google Earth が読める KML ファイルとして出力する関数や ESRI の ArcGIS が読めるシェープファイルとして出力する関数、面積を計算するプログラムなどを整備している。あらかじめ XML スキーマファイルから自動定義されるデータ型から Polygon2D への変換関数を定義し、JPGIS (GML) 形式の XML スキーマファイルを DPP に入力して建物外周線のデータを読み込ませたところ、正しくデータは読み込まれ、また、Polygon2D に対して定義された可視化関数の実行が可能であることも確認することができた。

5. まとめ

データの記述形式を自動変換することによりプログラムの適用範囲と再利用性を向上するデータ処理プラットフォーム (DPP) に対して、JPGIS (GML) 形式のデータの定義 (XML スキーマ言語で記述されたもの) を自動登録し、DPP に登録した JPGIS (GML) 形式のデータ間の変換を XML スキーマ言語で記述された継承関係に従って自動定義した。また、JPGIS (GML) 形式と種々の個別開発されたプログラムがもつ独自形式のデータとの間の変換を定義した。その結果として、異種データの連携が低コストで可能となった。

謝辞

本研究は、JSPS 科研費 JP18K13978、公益財団法人計算科学振興財団 研究教育拠点 (COE) 形成推進事業の助成を受けたものです。

参考文献

Muneo Hori, Tsuyoshi Ichimura, Lalith Wijerathne, Hideyuki Ohtani, Jiang Chen, Kohei Fujita, Hiroyuki Motoyama, 2018. Application of high performance computing to earthquake hazard and disaster estimation in urban area. *Frontiers in Built Environment*, 4, 1