

令和 2 年 6 月 23 日現在

機関番号：82401

研究種目：若手研究

研究期間：2018～2019

課題番号：18K13978

研究課題名（和文）記述形式の自動変換に基づく異種データ・異種プログラムの柔軟な連携技術の開発

研究課題名（英文）Data Abstraction and Data Interoperability Based on Automatic Conversion of Data Representations

研究代表者

大谷 英之（O-TANI, HIDEYUKI）

国立研究開発法人理化学研究所・計算科学研究センター・研究員

研究者番号：80639584

交付決定額（研究期間全体）：（直接経費） 1,600,000円

研究成果の概要（和文）：総合的な防災・減災の実現には、各分野の専門家が個別に開発する異種のプログラム群や関連する異種のデータ群を、柔軟に追加・取り替え可能な状態で連携させる方法論が不可欠である。従来、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が進められてきた。しかし、この方法論は、目的に適合した記述形式を設計して利用することを阻害する。そこで、本研究では、一律的な形式でデータを記述する標準化に替わり、記述形式を自動変換することでデータを抽象化する、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技術を開発した。

研究成果の学術的意義や社会的意義

本研究で開発した技術の特徴は、記述形式の自動変換によって独自形式のデータを接続していくことで、ボトムアップ的にデータの抽象化が可能な点にあり、この点がトップダウンで規定された形式に基づく標準化にない柔軟性を生み出す。これにより、複合災害の統合的な予測を行うシステムの持続的な拡張・高度化に対する大きな貢献が期待できる。また、一律的な形式でデータを記述する標準化を行っているあらゆる分野への応用が可能な技術であるので、総合防災のための異種データ・異種プログラム群の連携の促進に留まらず、多種多様な分野間の学際研究を一気に加速させる研究への進展が期待できる。

研究成果の概要（英文）：High-performance computing have enabled us to apply physical simulations to risk assessment for a variety of natural disasters, and recent data explosion can support this assessment approach providing the detailed information of the real world. However, the actual datasets, which is usually individually developed, are formatted differently even if the contents are equivalent, and it is required to start a digital city construction by making information extraction tools for individual datasets. It is a conventional solution to define a common standard format, but the solution cannot be applied to the existing miscellaneous formats. In this study, automatic conversion of data representations is proposed as a solution.

研究分野：防災工学

キーワード：総合防災 異種プログラム連携 異種データ連携 データ抽象化 疎結合

1. 研究開始当初の背景

現在、スーパーコンピュータ「京」の後継機が可能とする大規模数値解析に基づいて、地震・津波による複合災害を統合的に予測するシステムの開発が進められている(ポスト「京」重点課題3)。地震に対しては、地盤の震動から都市の個々の建造物の応答までが計算され、津波に対しては、地震や津波による建造物の破壊を考慮した浸水計算、人や車の避難計算が行われる。また、社会・経済への影響も計算される。開発には防災特有の多岐にわたる学際的な協力が不可欠であり、システム全体は、各分野内でタコツボ化した多数のプログラムが、都市の詳細や解析結果を表すデータの授受を通して結合されることで成立する。こうしたシステムの継続的な拡張・高度化には、システム内で連携するデータ・プログラムが容易に追加・入れ替え可能となる設計を当初から採用することが重要である。

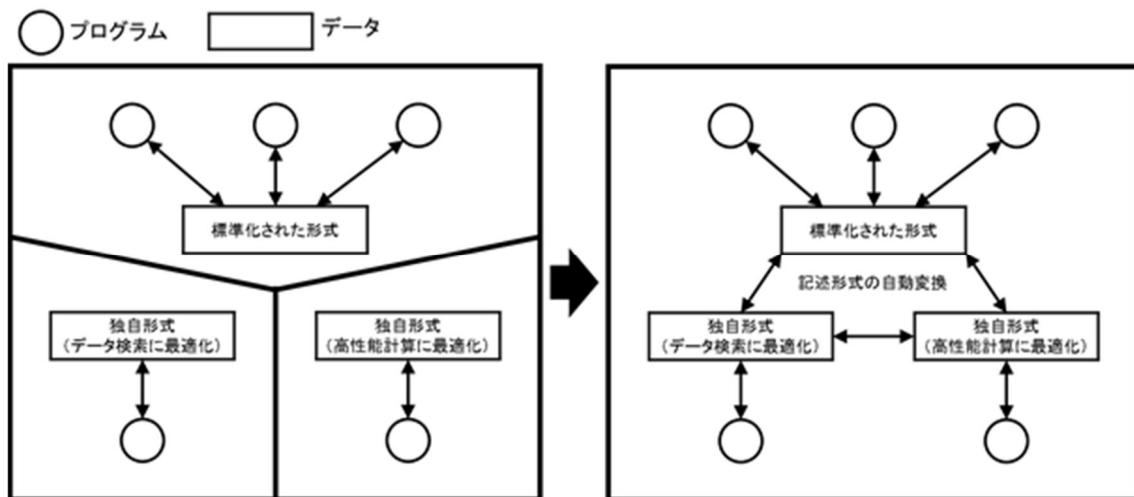


図1. 記述形式の自動変換に基づくデータ・プログラムのタコツボ化の解消

データ・プログラムが容易に追加・入れ替え可能な形で結合している状態は、疎結合と呼ばれるが、場当たりの開発で疎結合なシステムは開発できない。一般に、個別開発されたプログラムは個々の目的に最適化された独自の形式のデータを読み書きするため、プログラム間の結合には、プログラムの組み合わせの数だけのデータ変換プログラムが別途必要となるからである。この問題は、ややもすると単純な「記述形式の標準化」で解決可能と考えられる。すなわち、あらかじめ都市の情報や計算結果を多面的かつ詳細に記述できる汎用的なデータの記述形式を共通の認識に基づく標準として定めておき、すべてのプログラムにその標準的な形式に基づくデータを読み書きさせることで、プログラム間のデータの授受に変換プログラムが不要になると考えるのである。しかし、記述形式をトップダウンで一律的かつ固定的に定めるこの方法論は、データ提供者およびデータ利用者が、データ形式を独自に工夫してメモリ消費量を削減したりIO効率を改善したりすることを制限するものとなり、特に計算効率の最適化に特化したデータ形式が必須となる大規模数値解析プログラムには不適切である。

2. 研究の目的

本来、機械部品の規格などと異なり、データの記述形式は容易に変換可能なはずである。もし、図1の右のように、独自の形式と標準化された形式の間でデータを相互に自動変換できれば、独自形式のデータと標準化された形式のデータを同等に取り扱うことが可能となる。これは、標準化された形式のデータの代わりに、特定の目的に合わせて最適化した独自形式のデータをプログラム間のデータ授受に利用することが可能となることを意味し、このとき、自動変換で接続された同等なデータ群は、データの本質的でない差異を隠蔽して抽象化されたデータとみなすことができる。

本研究の目的は、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いたデータの抽象化によって疎結合を実現し、異種データ・異種プログラム群を柔軟に連携させる技術を開発することである。

3. 研究の方法

本研究の目的を達成するためには、

- (1) 記述形式の自動変換によってデータを抽象化する技術を開発することと、
- (2) データの抽象化による連携技術が一律的な形式でデータを記述する標準化による連携技術より有用であると示すこと

が重要である。

(1)の開発には次の ~ の達成が必要であるが、これらは高性能計算プログラムの入力となる都市のモデルを複数の情報源から統合して自動構築する別プロジェクトで既に達成済みである。

記述形式の自動変換に基づいたデータの抽象化を具体的に定義すること。

で定義されたデータの抽象化を実際に行うことができるようにすること。

で抽象化されたデータに対する処理を、コストが最小となる変換経路の自動選択に基づいて実行可能とすること。

の処理を簡易な記述で実行できるインタプリタを開発すること

これらの成果として開発されたデータ処理プラットフォーム(以降、DPP)は、地震・津波シミュレーションの都市モデルの構築に実践済みであり、独自形式のデータを取り扱うプログラム群を柔軟かつ効率的に連携させることが可能であることを確かめている。

(2)のためには、前者の枠組みに基づくDPPで後者の連携技術が取り扱えることを示し、前者が後者より包括的な方法論であることを示せばよい。

したがって、本研究の2年間では、一律的な形式でデータを記述する標準化の代表例として基盤地図情報で採用されている地理情報標準に基づくJPGIS(GML)形式を取り上げ、この形式で書かれたデータがDPPで取り扱えるようにする。具体的には、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DPPに登録し、また、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ことによって、個別開発されたために連携が難しかったプログラム群を、JPGIS(GML)形式で書かれたデータを介して連携可能とする。

技術的な課題は、XMLスキーマ言語などで記述されたデータの定義の自動登録を可能とすることである。現状のDPPで取り扱えるデータの種類の10のオーダに留まっている。これは、取り扱えるデータを手作業で登録して増やしているからであり、様々な標準においてXMLスキーマ言語などで定義された100~1000のオーダの種類のデータに対応することが事実上不可能となっている。

この課題を解決するため、平成30年度には、

- ・XMLスキーマ言語で定義されたデータをDPPに自動登録する機能を開発し、
- ・JPGIS(GML)形式について自動登録の検証を行う。

そして、平成31年度には、

・独自形式との間の自動変換を定義してJPGIS(GML)形式で書かれたデータを介した個別開発されたプログラム間の連携を可能とし、

・また、JPGIS(GML)形式のバージョンの違いについても、この差をデータの抽象化によって隠蔽することが可能であるため、容易に対応可能であることを示すことで本研究の目的を達成する。

4. 研究成果

(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) 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) 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。

(4) 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。

(5) 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。

(6) 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。

(7) 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 スキーマファイルで 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。

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。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 大谷 英之
2. 発表標題 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
3. 学会等名 第28回地理情報システム学会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----