

XML を利用した地球惑星科学データベースの考え方

XML database application for earth and planetary sciences

寺園 淳也[1], 齋藤 潤[2]

Jun-ya Terazono[1], Jun Saito[2]

[1] (財)日本宇宙フォーラム, [2] 西松建設(株)技術研究所

[1] JSF, [2] Technical Research Inst., NISHIMATSU Construction Co., Ltd.

<http://www.t3.rim.or.jp/~terakin/>

惑星探査の進展やデータの高精度化により、我々が扱っている科学データの量はテラバイトクラスに達しようとしている。このような大容量の科学データを有効に活用するため、我々は DPLEX(Desktop Lunar Exploration)という環境を提唱している。これは、ネットワークで結ばれた科学データサーバとネットワークで結ばれた科学者や個人が自由に惑星探査データを解析できる環境である。この前提となる科学データの相互交換に関し、現在ネットワーク技術で注目されているXMLが応用できるかどうかを考える。特に、既存のデータフォーマットとの相互運用性や、現在実装が進んでいる天文分野の例などについて紹介する。

惑星探査の進展や大型観測計画、さらには科学データそのものの高精度化により、我々が扱っている科学データの量は指数関数的に増大している。これから数年後の惑星探査ミッション、及び大規模な科学観測などで発生するデータ量は、年数十GBクラス、全体では数TBクラスに達すると考えてよい。

一方で、こういったデータ解析を行う科学者や技術者の側の体制整備はまだ、ようやく始まったばかりである。ほとんどの研究者は、依然として過去から受け継がれた解析プログラムを自分たちで組み直して、解析を行っているのが現状であり、早晚その非効率性が限界に達することが考えられる。貴重な国民の税金を投入して得られた惑星探査データを無駄にすることは許されない。そのためには、データを有効に使う枠組みが必要である。

膨大な探査データから科学者や技術者が適切なデータを選び出せるようにするためには、

- ・処理すべきデータに的確にたどりつけるようにするためのシステムの開発
- ・大量のデータをいかにわかりやすく可視化するか、そのためのシステムの開発

が必須である。また、これらの技術は、こういった大規模科学データを一般の人々に伝える際にも応用が可能である。

このような大規模科学データの解析の枠組みとして、我々は Desktop Lunar and Planetary Exploration (DPLEX)という概念を提案している。これは、ネットワークで結ばれた科学データサーバと、過去の探査データや科学情報などを蓄積したサーバなどを利用して、ネットワークで結ばれた科学者や個人が自由に(自分のコンピュータのデスクトップ上で)惑星探査データを解析できる環境である。

ここでは、この DPLEX システムを実現していく上で必要となる、個別の要素技術について述べる。それにより、どのようにして DPLEX を実現させていくかについての見通しについて議論する。

ネットワークを通じて大規模科学データをやりとりする場合には、多様なデータフォーマットを統一したやり方で扱えるような技術が必要になる。このような情報流通基盤として、現在最も注目されている技術として、XML (eXtensible Markup Language)がある。

XML は、SGML をもとに、現在のネットワークにおけるデータ流通に必要な機能を盛り込んだデータ表現言語である。XML の特徴は、次の通りである。

- ・データスキームを定義する言語であり、拡張性が最初から前提となっている。例えば、タグを組み合わせて新しいタグを定義したり、新しい処理スキームを宣言することも可能である。
- ・既に標準化が進められており、プログラム開発を行いやすい。既に様々なシステムや言語に対応したXMLパーサが開発されているため、これらを利用して、科学データ処理のためのシステムを構築することができる。
- ・出力形態としてHTMLが前提となっているため、既存のインターネット技術と親和性が高い。
- ・3D表現やベクトル表現、数式など、科学技術データの可視化に応用が可能な技術が既に標準化されつつある。

XMLを利用した科学データの取り扱いに際しては、そのデータの取り扱いを定めたファイルである、文書型定義(DTD)を作っていく必要がある。さらに、それが正式な規格として採用されるように、科学分野ごとに調整を図らなければならない。

DTD開発の動きは天文分野を中心に進められている。既に、NASA GSFCでは、Astronomical Data Center XMLとして、天文データのXML化に必要なDTDなどが公開されている。さらに、イリノイ大学のD. Guillaume氏によっ

て、天文観測データ及びそれに付随するメタデータを記述するための XML 言語規格、AML (Astronomical Markup Language) が提案されており、簡単なブラウザ・アプレットも公開されている。

XML 自体は標準規格であるが、その名称が示すように極めて拡張が容易な言語である。従って、これから科学データに関する多くの XML 規格が提唱され、互いに競争しながら生き残っていくと考えられる。規格を提案してことはもちろん必要であるが、データごとの標準化プロセスや、たとえば天文と地球物理などで共用できる DTD などを開発していく必要もあると考えられる。

本講演では、この XML を将来の科学データの標準交換プラットフォームとするために、特に地球惑星分野でどのような工夫が必要であるかについて、その展望について述べる。同時に、天文をはじめとした関連分野における XML を用いた科学データの取り扱いについての最新の情報を紹介する。