

日本の月惑星探査と科学データアーカイブ (第1回 PDS と日本の現状)

山本 幸生 (宇宙科学情報解析研究系)

はじめに

日本の惑星科学分野においてデータアーカイブは今まさに黎明期と言っても過言ではありません。これは小惑星探査機「はやぶさ」や月周回衛星「かぐや」が、日本として公開可能なまとまったデータを取得したこと、また将来構想としてシリーズ化した探査計画の具体的なビジョンが示されたことに起因しています。これまでのプロジェクト独自にデータアーカイブしていた状況から、プロジェクトを超えて、同種の要求に対して同種のサービスを提供可能な、全体としての枠組みを考慮する必要性が生じています。今回の PLAIN News から不定期に、日本の月惑星探査と科学データアーカイブの連載を開催していく予定です。

Planetary Data System

惑星科学の分野では、事実上の標準として、NASA の開発した Planetary Data System(PDS)*¹ が最もよく利用されています。というのも、現在惑星探査によるデータは NASA が最も多く保有しているからです。PDS は探査プロジェクト終了後の長期保存・利用可能性を主眼として高度に設計されています。PDS の詳細については NASA PDS のウェブサイト²で参照可能です。他の分野で異なった意味で使っていることも想定して、最初に PDS で使われる用語について整理します。

PDS では画像やヒストグラムなどの観測機器等が出力するデータのことを「(データ)オブジェクト」と呼び、それらに付随する情報(機器名称や観測時刻など)を「ラベル」と呼んでいます。オブジェクトやラベルは「Primary Data(主要データ)」と呼ばれ、詳細説明の書かれたドキュメントやソフトウェア、全体の構成が記述されたカタログなどを「Ancillary Data(補助データ)」としています。

科学的に意味のあるデータの最小単位を「(データ)プロダクト」と表現し、全てのプロダクトは「Primary Data」と「Ancillary Data」から構成され、通常複数のファイルで1つのプロダクトを成しています(図1)。1つのプロダクトには画像やヒストグラムといった複数のオブジェクトを含めることが可能となってい

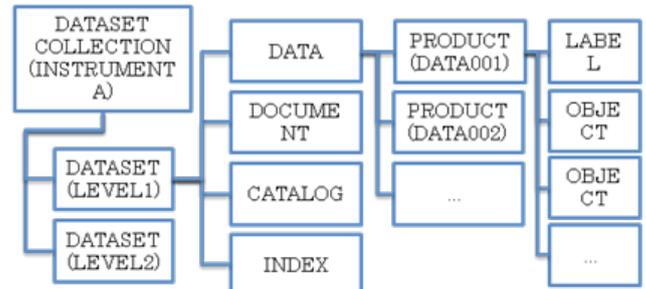


図1 PDS のデータ構造例

ますので、例えば3種類のフィルタで同時刻に取得した画像などは1つのプロダクトにまとめることができます。一般的にデータ解析を行うユーザは、「プロダクト」単位でウェブなどから取得し、プロダクトに含まれるオブジェクトを解析することになります。

ラベルとオブジェクトはしばしば「Attached 形式」として1つのファイルにまとめられますので、このファイルだけを見ると、天文学の分野で頻繁に利用される FITS フォーマットや、流体分野で利用される CDF 等と良く似た「メタ情報を保有したデータフォーマット」として認識されます。

またプロダクトの集合のことを「データセット」と呼び、観測機器ごとの違いや処理レベルの異なるデータについてはデータセットで区別します。さらにデータセットの集合のことを「データセットコレクション」と定義し、同一の観測機器は同一のデータセットコレクションとするのが通例となっています。

PDS 自身は任意のフォーマットを入れるための「入れ物」として機能している面もあり、PDS の中に別のフォーマット(例えば JPEG や PDF)でデータを格納することも可能となっています。その一方でラベルについては、例えば地理情報に関しては Cartographic Standard に従ったキーワードを使用したり、衛星の姿勢や軌道に関するキーワードを予約語としたりするなど、惑星科学や探査に特化した一面もあります。

他にも CD や DVD などのメディアで配布することを想定したインデックスや、ミッションや観測機器、データの全容を記述したカタログファイルも PDS の仕様として盛り込まれています。

[裏へ続く]

このように、PDS はフォーマットと言うよりはその名の通りシステムであり、単一のファイルフォーマットのみを規定している訳ではありません。PDS に従ってデータ整備をすることは、NASA が蓄積してきたノウハウをそのまま流用することを意味しており、後発の国々がお手本とするには非常によい教材となっています。

PDS の問題点

PDS の体系は利用者にとって非常に有用ですが、PDS に正式に準拠していることを謳うには、NASA による厳密な審査過程を必須としています。そのため米国以外の国が PDS を整備することを考えた場合、各国が独自で取得したデータであるにも関わらず主導権が発揮しにくいデメリットがあります。

また PDS はその規模が膨大であるために、プロジェクトごとに PDS を習得するにはオーバーヘッドが大きすぎる面もあります。

これに対して ESA では、PDS のラベルとデータ本体であるオブジェクトのみ同一フォーマットを採用し、ドキュメントやカタログ等については独自で整備する方式を採用して Planetary Scientific Archive (PSA)*² としてまとめています。

JAXA も基本的に ESA と同じ方針を採用し、特別な名称は今のところありませんが、PDS ライクや PDS 準拠と呼んでデータを整備しています。それでも小惑星探査機「はやぶさ」のように、プロジェクトごとに NASA との協力関係を結び、可能な範囲で NASA によるレビューを受け、正規の PDS として公開を行っている部分もあります。

PDS の別の問題点として、PDS は任意のフォーマットを格納できる点があります。これは何を意味しているかということ、フォーマットを決めているようで何も決めていないこととなります。惑星探査の多様な観測機器のデータを格納可能な柔軟性を求めた結果、データと直結したフォーマットが定まらず、利用可能なアプリケーションが育ちにくい状況にあります。PDS でデータを格納する場合には、慣れ親しんだソフトウェアで解析するために、もう一段掘り下げて、内部的な格納方法について検討する必要があります。

日本の月惑星科学分野の問題点

現在 JAXA で取得した惑星探査・科学衛星データ

を長期保存および公開していくために、PDS のノウハウを十分に吸収して、さらに日本の開発状況に馴染んだ手法でデータアーカイブを行なうことを考えています。

日本の科学衛星の特徴として、科学者自ら観測機器を開発・校正・試験を行い、さらには打ち上げ・衛星運用・データ解析・データアーカイブ・公開準備、これら全てを観測機器チームが主体となって行う一極集中型の開発が挙げられます。

全体像がよく見えるという意味では良いのですが、データアーカイブ・公開準備に関しては軽視されている面も否めません。というのも日本の宇宙開発の科学者は複数のプロジェクトに関与していることがほとんどで、いざデータアーカイブの段階に差し掛かると、既に別のプロジェクトの開発で忙しく余裕がない、という状況に陥っていることがしばしばあります。そのため第三者が利用可能なように、PDS と同品質でデータが整備されることはほとんどの場合ありません。

月惑星科学分野におけるデータアーカイブの枠組み構築

科学者とは別に、PDS の膨大なノウハウを効率よく吸収し水平展開するために、データアーカイブ専門の部署が必須であると考えています。そういう意味で「宇宙科学情報解析研究系」や「科学衛星運用・データ利用センター (Center for Science satellite Operation and Data Archive; C-SODA)」、あるいは「惑星探査グループ (JAXA Space Exploration Center; JSPEC)」内に専門部署を設けるなど、幾つか候補が考えられます。

部署としては候補がありますが、肝心のノウハウはまだ吸収されていないため、まず PDS やその他のノウハウを吸収する人材の確保と、今後全ての惑星探査プロジェクトに適用可能なように、衛星開発プロジェクトに早い段階で組み込むスキーマ作りを行う必要があります。科学者の負担を低減しながらも、取得したデータを高い品質で保存・公開することを実現する環境を構築していければと思います。

参考ウェブサイト URL

*1. NASA PDS <http://pds.jpl.nasa.gov/>

*2. ESA PSA

<http://www.rssd.esa.int/index.php?project=PSA>

編集発行：宇宙航空研究開発機構 宇宙科学研究本部 科学衛星運用・データ利用センター
〒229-8510 相模原市由野台 3-1-1 Tel.042-759-8767 住所変更等 e-mail : news@plain.isas.jaxa.jp
本ニュースはインターネットでもご覧になれます .<http://www.isas.jaxa.jp/docs/PLAINnews>

●編集後記：ハーブの発表会でホルストの「ジュピター」を弾きました。C-SODA の同僚が聴きにきてくれたのに、緊張してうまく弾けなくて残念！深夜のオフィスで一人で練習しているときは完璧に弾けるのに…(K.E.)