

惑星探査データの保存と公開 ～ PDS4の紹介～

山本 幸生¹

1. はじめに

室内実験と比較して、惑星探査による「その場観測」は金銭的成本が高く、その結果データを得る機会は制限され、同種のデータを取得することが容易でない。科学における客観性と再現性という意味において、この容易に取得することのできないデータを第三者が利用できるように整備することは、惑星探査の科学的価値を維持する上でも重要な作業であると言える。

こと惑星探査においては、データを保存するための仕組みとして、NASAが開発しているPlanetary Data System(PDS)と呼ばれるシステムが事実上の標準(de facto standard)として利用されている[1]。De factoとして採用されている理由は単純で、NASAが最も多くの惑星探査機を打ち上げ、最も多くの惑星探査データをアーカイブしているからに他ならない。本稿では惑星探査データの特徴に触れながら、このPDSの最新状況について紹介する。

2. 惑星探査データの特徴とPDS Nodes

一重に惑星探査データと言っても単純ではない。惑星探査機には様々な学術分野の観測機器が搭載されており、各々の学術分野には歴史ある「しきたり」が存在する。研究者にとって「しきたり」に従った方法でデータがアーカイブされていることは、データを扱った研究を始める上での敷居を下げることになる。PDSでは分野横断的なデータアーカイブを実現するために、PDS Nodesと呼ばれる学術分野や技術分野のまとまりを切り口とした拠点を設け、そこからデータのアー

カイブとサービスを提供している。

3. PDSの意味

PDSという言葉は場面によって異なる意味で使用される。「PDSに問い合わせる」という使い方は、NASAの一部門としてのPDS部門やPDS Nodesの個々のノードを意味している。また「PDSからデータを公開する」では、PDS部門やPDS Nodesが個々に管理しているウェブサイトのことを意味する。本稿で紹介する「PDSでデータをアーカイブする」という文脈におけるPDSの使い方は、「データアーカイブの技術的な仕組み」のことを指すが、難解なことに、この表現はしばしば混乱の元となる。それは「何をどこまで行ったらPDSでデータをアーカイブしたことになるのか」という点を明確に示していないため、受取手によっては異なった対応を想像するからである。

4. PDSの対応レベル

曖昧さを減らすために、PDSの対応レベルとして4段階を定義する。

- (1) PDSに従ったフォーマット・ディレクトリ構造でファイル群を保存する
 - (2) PDSが求めるドキュメント整備を行う
 - (3) PDSのピアレビュー(査読)に合格する
 - (4) PDSのウェブサイトからデータを公開する
- (1)と(2)は技術的・人的リソースを含んだコストの問題であり、(3)と(4)宇宙機関をまたぐ政治的な問題と言える。ESAでは(1)と(2)のみ対応したPlanetary Sciece Archive(PSA)を開発し、ピアレビューやデータの公開は独自に行っている。JAXAも基本ESA

1. 宇宙航空研究開発機構 宇宙科学研究所
yukio@planeta.sci.isas.jaxa.jp

と同じ方針としているが、プロジェクトによってはPDSからデータを公開することもある。

5. PDSウェブサイトから公開することの課題

技術的な話に入る前に、ここでPDSのウェブサイトから公開することの課題を述べておく。PDSのウェブサイトには、世界中の惑星探査データをアクセスする人が集まるため、データを本当に利用してもらいたいと望むのであれば、PDSのウェブサイトからデータを公開する意義は大きい。ではなぜJAXAはPDSからデータを公開することを躊躇するのか、それは主導権と知的財産権に関係しており、そのことについてJAXA内で十分な議論がなされていないためである。

PDSのウェブサイトからデータを公開するためには、論文と同様に「ピア・レビュー」と呼ばれる査読プロセスを通さなければならない。このことはすなわち、日本のデータで日本が公開したいと希望するのに、PDSに投稿しPDSのウェブサイトから公開する以上、査読結果によってはリジェクトされ、データの主導権がNASAに握られてしまう可能性がある。

またPDSのウェブサイトからデータを公開すると、特殊な場合を除き、それらのデータはパブリックドメイン扱いにされてしまう。パブリックドメインとは、知的財産権がないことを意味しており、商用利用や改変に関する権利を放棄することになるため、今のところJAXAの方針と相容れるものではない。

6. PDS4概要

PDSの開発は1990年のPDS1からスタートした。現在は2010年に公開され今もなお開発の続いているPDS4が最新版である。PDS4はIT分野の発展を十分に取り入れた設計が行われている。メタデータ記法、データ型の再定義、用語の再整理が行われ、これまでのPDS3とは大きく変わっている。

PDS3は、敢えて言うならば、天文分野で良く使われるFITS形式[2]に似通っている。観測年月日や観測領域などのメタ情報部分は、FITSヘッダーの代わりにPDSラベルと呼ばれるテキストファイルを利用し、そこにObject Description Language(ODL)という独

自の記述形式を用いる。特にPDSラベルとデータ部分を1つのファイルに合わせ込むDetached形式は、他のファイルフォーマットを強く意識していることが感じられる。

PDS4では、メタ情報部分はXMLという汎用な記述方式に置き換わり、設計にはXMLスキーマ(拡張子: xsd)を利用し、スキマトロン(拡張子: sch)を使って検証ルールを作成する方式へと変更された。この一般的な技術の導入により、周辺システムとの親和性が向上し、また関連技術を有する技術者の調達が容易となった。しかしながら、惑星科学を主たる専門とする研究者が、PDS4に準拠したデータアーカイブを自力で行うのは、もはや非現実的な段階へと差し掛かった可能性がある。

PDS4のデータ部分は、4つの基本的なクラスへと再整理された。配列型(Array)、テーブル型(Table_Base)、構文解析可能なバイト列(Parsable_Byte_Stream)、符号化バイト列(Encoded_Byte_Stream)である。この再整理により、既存フォーマットとの対応がより明確になった。例えばシンプルなFITSフォーマット(Binary Table Extensionを利用していないFITS)の画像データはPDS4のArrayと互換性がある。固定長のASCIIで書かれたテーブルはTable_Baseとして、HTML、XML、CSVなどはParsable_Byte_Stream、JPEG、PNG、PDFはEncoded_Byte_Streamといった具合である。

したがって既存のフォーマットを流用すれば、データ部分の準備は、比較的少ない手間でPDS4に対応することが可能である。しかしながら、PDS4のデータ部分はEncoded_Byte_Streamを用いるべきでないというポリシーがある[4]。これはJPEGやPNGによるデータの公開はPDSとして認めないことを案に表明しているが、その一方で同様に複雑なデータ構造を持つCommon Data Format(CDF)も除外されてしまうなど、課題点は残っている。

7. PDS4によるアーカイブ例

PDS4には論理的な概念としてBundle-Collection-Productの層構造がある。これはファイルシステムを考えると比較的容易に理解可能である[3]。Bundleはトップディレクトリ、Collectionはその下に作られる

```

HAY2_ONC
├── bundle.xml
├── context
│   ├── collection.tab
│   ├── collection.xml
│   ├── instrument_ONC_1.0.xml
│   ├── instrument_host_HAYABUSA2_1.0.xml
│   ├── investigation_HAYABUSA2_1.0.xml
│   ├── node_SBN_1.0.xml
│   ├── other_dataset_1.0.xml
│   ├── other_reference_1.0.xml
│   └── target_1999ju3_1.0.xml
├── data
│   ├── 20181001
│   │   ├── abc.fit
│   │   └── abc.xml
│   ├── collection.tab
│   └── collection.xml
├── document
│   ├── alignment_stereo
│   │   ├── alignment_stereo.pdf
│   │   └── alignment_stereo.xml
│   ├── bibliographic_information
│   │   ├── ads.asc
│   │   └── ads.xml
│   ├── collection.tab
│   ├── collection.xml
│   └── observation_log
│       ├── oncobslog.asc
│       └── oncobslog.xml
├── readme.txt
└── xml_schema
    ├── PDS4Lb1XML_SDU_Instrument_Product_Information.xsd
    ├── PDS4_PDS_1101.sch
    ├── PDS4_PDS_1101.xml
    ├── PDS4_PDS_1101.xsd
    ├── PDS4_hay2_dict_1101.xsd
    ├── collection.tab
    └── collection.xml

```

図1：PDS4のディレクトリ構造例. 黒字(太字)は観測機器チームで用意. 赤字はPDS4のウェブサイトからダウンロードするファイル.

ディレクトリ (browse, context, data, document, xml_schema, etc.), またProductは各種Collectionディレクトリに含まれるファイルに相当する.

図1に「はやぶさ2」のカメラ(ONC)を想定したPDS4のディレクトリ構造例を示す. この例ではHAY2_ONCと言うBundle名称に, context, data, document, xml_schemaの各Collectionが用意されている. プロジェクト用意すべきファイル, 観測機器チームが用意すべきファイルが黒字で示されている. この例における実際のデータ本体は, FITS形式のabc.fitと, そのメタ情報を含むabc.xmlである. データ本

体以外の部分で用意すべきファイルが多いことが分かる. これこそがPDS4でデータを整備することの, すなわち長期的なアーカイブをめざした場合に必要な手間なのである.

8. おわりに

NASAは2011年11月以降打ち上げの全ミッションに対してPDS4を適用することを決めた. また惑星探査データの標準化に対して意見の言える国際惑星データ連合(International Planetary Data Alliance; IPDA)

[4]は、PDS4を惑星探査データアーカイブの標準として推奨した。このことは即ち、PDS3に関するNASAのサポートは絶望的であり、一方でPDS4を使用するのであれば全面的な協力が得られる状況にある。日本は継続中のミッション(あかつき)ではPDS3を採用し、今後打ち上げられるミッション(はやぶさ2, BepiColombo)についてはPDS4を採用する予定である。そのため一刻も早くPDS4を構築する体制を作らなければならない。

参考文献

- [1] The Planetary Data System, <http://pds.nasa.gov/>.
- [2] Wells, D.C. et al., 1981, Astron. Astrophys. Suppl. 44, 363-370.
- [3] PDS4 Concepts Version 1.0.0, http://pds.nasa.gov/pds4/doc/concepts/Concepts_130507_v1.pdf.
- [4] Data Provider's Handbook, http://pds.nasa.gov/pds4/doc/dph/current/PDS4_DataProvidersHandbook_20130507.pdf.
- [5] International Planetary Data Alliance, <http://planetarydata.org>.