

```
PDS_VERSION_ID = PDS3
```

```
/* FILE P
```

```
RECORD_TYPE  
RECORD_BYTES  
FILE_RECORDS  
LABEL_RECORDS  
INTERCHANGE_FORMAT
```

メタデータを考慮した 地球惑星科学 データフォーマット

```
/* POINTERS TO START RECORDS OF OBJECTS IN FILE */
```

```
^IMAGE
```

```
/* IMAGE I
```

寺菌 淳也 Junya Terazono
会津大学
terazono@u-aizu.ac.jp

```
DATA_SET_ID = CLEMT-L-0-5-DIM-UVVIS-V1.0  
PRODUCT_ID = "UI10S027"  
PRODUCER_INSTITUTION_NAME = "UNITED STATES GEOLOGICAL SURVEY"  
PRODUCT_TYPE = MDIM  
MISSION_NAME = "DEEP SPACE PROGRAM SCIENCE EXPERIMENT"  
SPACECRAFT_NAME = "CLEMENTINE 1"  
INSTRUMENT_NAME = "ULTRAVIOLET/VISIBLE CAMERA"
```

```
ui10s027.img lines 1-23/77418 0%
```



背景

- 近年、地球惑星科学に限らず、あらゆる領域で「メタデータ」の重要性が高まりつつある。
- 科学観測においては、観測データそのものではなく、それに付随するデータ、すなわち「メタデータ」の存在が重要である。
- 地球惑星科学においても、こういったメタデータの存在は重要である。
- しかし、現在の地球惑星科学(特に惑星科学)のデータフォーマットでは、こういったメタデータへの対応は不十分である。
- そこで、本発表では、メタデータに十分対応した新たなデータフォーマットについて考察し、その可能性について検証したい。



メタデータとは

- メタデータとは、データに付随するものである。
 - メタ(抽象的な上位)データである。
 - それ自身もデータの形をとっているから、基本的にはデータと見なされる。
- データに付随するものであるにもかかわらず、データとは別に存在するものである。
 - 例えば、本をデータとするならば、著作者や表題のような、データを抽象化するものであるがそのデータを表すのに役立つ概念を指している。
- 「メタデータ」には、データに直接付随するものと、コメント的なより広い意味を持つものとの2つがある。



惑星科学データの基本

- 多くの惑星科学データは、「ヘッダ＋データ」の形で表される。
 - 惑星画像をはじめとする惑星探査データの格納形式のPDS、天文データのFITS、流体力学データのNetCDFなど、ほとんどのデータそうといってよい。
- ヘッダとして表現されたデータは、本体のデータの要約と、本体のデータからは再現されないデータの2種類に分かれる。
 - 本体のデータの要約は、迅速なデータ解析や、データそのものの要素の表現のために便宜的につけられるケースが多い(例えばデータの平均値など)。
 - 問題は、本体から再現されないデータ。これが、本来の意味のメタデータ。
- これらのデータは、できうる限り一体化していることが望ましい(1ファイルではなく、リンクしていればそれでもよい)。



データは何が 保証されるべきか

- アクセシビリティ

- 大容量データの解析時代、科学者はなるべく簡便な手段でデータアクセスを行うことにより、手間を最低限に減らしていくことが必要。
- データが分散して保存されている場合、データ間の連携がとれることが重要。

- 永続性

- いつのデータであっても、そのデータが得られた状況などをしっかりと把握できることが必要。
- フォーマットの仕様自体も何らかの形で残っていなければならない。

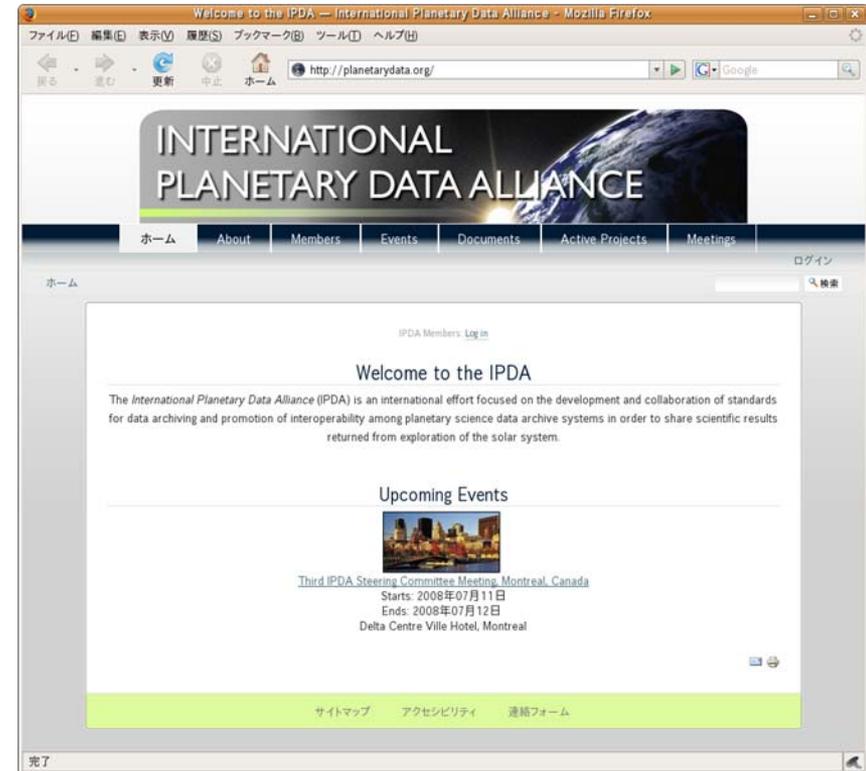
- データの説明

- そのデータが何を意味しているのかが、明確な形で理解されなければならない。



XMLベースの 惑星科学データフォーマット

- 現在、ヨーロッパを中心とした科学チーム(IPDA: International Planetary Data Alliance)などが中心となり、PDSに代わる新たな惑星科学データフォーマットの策定作業を進めている。
- 現在のところ仕様はまだ0.3版と、非常に暫定的なバージョン。
- 基本的にXMLベースになる模様。



IPDAのウェブサイト
(<http://planetarydata.org>)



XMLをベースとした 一体型のデータファイル構造

- この場合には、メタデータも含めてすべてを1つのファイルの中に記述する方式になる。
- ミッションの種類やバージョンの違いはDTDで吸収する。
- 実データをどのように入れるか(バイナリにするのかテキストにするのか)がポイント。バイナリXMLの活用がキーポイントか？

```
<?xml ...>  
<xmlns...>
```

```
<description>  
データ記述  
</description>
```

```
<data>  
....  
</data>
```

メタデータ記述部。旧来の「ヘッダ」を置き換える。RDFなど。

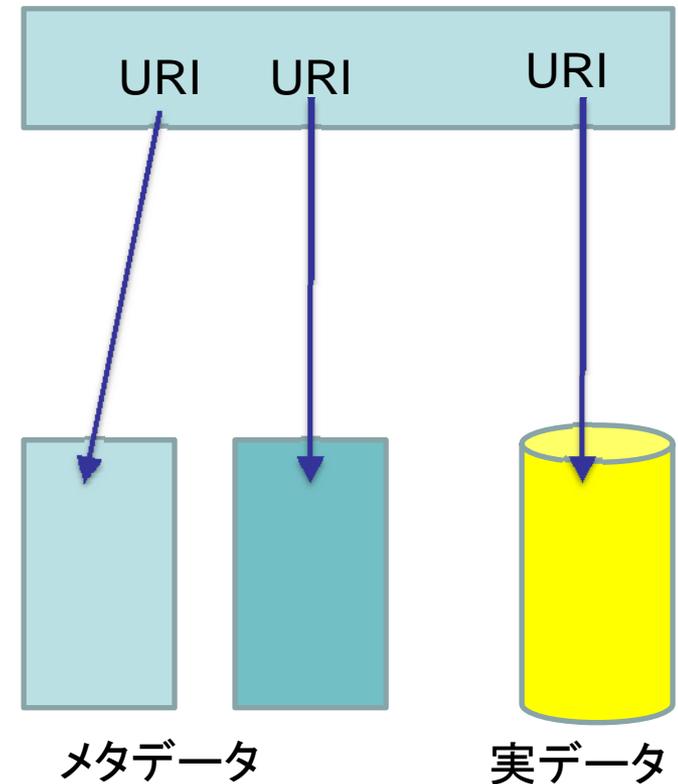
データ記述部。テキストまたはバイナリ。ただ、混在できるかどうか。



XMLをベースとした 分散型データ構造

- この場合、データは1つのファイルの形ではなく、いくつかのファイルまたはロケーションに分かれたデータの形をとる。
- それぞれのデータはURIによって結合される。
- URI自体の永続性が保証されなければならないが、仕様定義等は個別のデータに対して行えるのでかなり楽ではある。またオブジェクト指向的であるともいえる。
- クラウドコンピューティングなどにも親和的。

データファイルとしての実体
(主体)





データ構造とは？

- これまでは、データ構造は「ファイル」という形でとらえられてきた。従って、ファイルフォーマットがデータ構造の大きなポイントになっていた。
- しかし、インターネットの時代になり、ネットワークで結ばれた分散コンピューティング時代に入りつつある(あるいは将来そうなる)という状況においては、ファイル構造よりはデータのロケーションが重要ということになる。
- 将来的には、後者の分散型のデータ構造(各データはXMLで記述、ないしはやりとりされる)という構造になり、データそのものすらURIで記述される時代になるのではないか(XMLは単なる要素技術としての通過点となる)。



より科学者に使いやすい データ構造を目指して

- こういったデータ構造は、使いやすくなければ科学者にとっては意味がない。
- アクセスするためのツールや、仕様を一意的に決めることができる(決まる)仕様書などが必ず必要。
- URIの永続性が保証されていない現状では、それを担保するようなインターネット技術の発展も必要になる。
- XMLはあくまでも、データのやりとりの要素技術として考える必要がある。データがテキストであろうとバイナリであろうと、あるいはデータベースに入っていようとテキストファイルに記述されていようと、一意にアクセスできることが重要。
- 今後IPDAのフォーマット進展をみながら、さらに次世代のフォーマットを考えていきたい。