

宇宙科学データベース DARTS

Data ARchive and Transfer System*

田村 隆幸, 馬場 肇, 松崎 恵一, 三浦 昭, 篠原 育,
長瀬 文昭, 笠羽 康正, 上田 佳宏 (JAXA/ISAS)

1 DARTSの目的

DARTS¹は宇宙研・科学衛星の観測データを、多くの研究者が広く活用できるように、宇宙科学情報解析センターが中心となって構築したデータベースシステムである。公開データを保存・配布し、利用者に以下のサービスを提供する。

- (1) データ検索、
- (2) データのオンライン解析環境、および
- (3) データおよびデータ解析に関する情報。

世界中の研究者がこれらのデータに、「高速に」、「高い信頼性で」、かつ「容易に」アクセスでき、「柔軟な」解析ができるシステムを目指している。

2 現在のサービス

現在、表 ??にまとめた宇宙研のデータを DARTS から公開している。宇宙研内外のプロジェクト関係者と共同でこれらのコンテンツを整備してきた。これらに加え、いくつかの海外データのミラーリングを提供している。

3 データ検索・転送 Web システムのソフト開発

2004 年度より (株) セックと共同でデータ検索・転送システムを開発している。主目的は、2005 年以降に打ち上げられる大型プロジェクト Astro-

E2、Astro-F、Solar-B のデータ公開に対応するためである。

これらの衛星のデータ量は、これまでのものに比べざっと 10 倍程度になり、同時に DARTS を利用するユーザの数もぐっと増える。既存の DARTS では、複数の検索・データ管理システムがそれぞれの分野 (天文、太陽観測、太陽地球物理)、ときには各衛星ごとに混在している。また、それぞれのシステムも数年前に設計されたもので、新規のプロジェクトへの拡張には大きな手間がかかる。そこで、これまでの我々の開発経験と現在の新規技術を用いて、新たなデータ検索・転送システムを開発することとした。

設計を始めるにあたり、既存のシステムの問題点を洗い出した。現在の DARTS 天文分野のシステムは、

- (1) CGI/Perl によるユーザ I/F、
 - (2) C 言語による検索を処理するソフト、
 - (3) リレーショナルデータベース管理システム (Oracle9)、
 - (4) オリジナルのデータカタログ、
- およびこれらの間の I/F からなっている。この構成は、以下の点で、移植性、拡張性が高くない。
- (A) C 言語に依存した部分がある、
 - (B) (2), (3) の間の I/F は、Oracle 固有のライブラリーを用いている、
 - (C) (4) を (3) に入力する部分に手作業が必要。

また、CGI/Perl によるユーザ I/F は、複数のリクエストを同時に処理するのに向いていない。

優先した開発コンセプトは以下のとおりである。

信頼性 基本機能をユーザに確実に提供する。世

*宇宙科学シンポジウム 2005-1 @ ISAS

¹www.darts.isas.jaxa.jp

表 1: DARTS で公開している JAXA/ISAS の科学データ

プロジェクト#1	FTP#2	HTTP#3	データの種類
X線天文			
ASCA ^b (1993/2-2001/3) (340 GB)	Yes	データ検索 CGI	テレメトリデータ 1次処理データ 較正データ QLデータ
Ginga (1987/2-1991/11) (34 GB)	Yes	ASCA と共通	テレメトリデータ (LAC) 解析ソフト
Tenma (1983/2-1989/1)	NO		観測ログ
電波天文			
HALCA ^b (1997/2-)	Yes	ASCA と共通	公開用データ
赤外線天文			
IRTS (1995/3-1995/4) (2GB)	Yes	データ検索 CGI	点源カタログ 輝度イメージ
太陽観測			
Yohkoh (1991/8-) (5GB)	Yes	データ検索 Java	テレメトリデータ 1次処理データ 観測ログ 日にちごとのイメージ
太陽・地球物理			
Geotail (1992-) (8 GB)	NO	データ検索 CGI	磁場 (MGF) 低エネルギー粒子観測器 (LEP) プラズマ波動 (PWI)
Akebono (1989/2-) (25 GB)	NO	データ検索 CGI	軌道・機器情報 低エネルギー粒子観測器 (LEP) オーロラ撮像器 (ATV) 放射線モニター (RDM)

#1: プロジェクトの期間と DARTS で公開しているデータ総量 (2005/1 現在), #2: FTP でのデータ公開の有無, #3: HTTP で提供しているサービスとその実装言語。

界中のユーザから利用されるので、可用性が高い(障害によるシステムダウン時間の短い)ことが必須。

適時性 新規のプロジェクトのデータ公開までには確実に試験を終わり、すぐに運用に入れるように開発スケジュールを立てた。

拡張性と管理の容易性 DARTSでは分野の異なる複数の衛星データを扱う。また、個々の衛星データにおいても、運用期間中に、データベースの項目が変わる可能性がある。そのような拡張に柔軟にかつ容易に対応できる構成にする。

このような設計を実装するため、CGI/Perlに替え、Javaによるコード化を採用した(図1)。また、システム内の情報管理および外部システムとのI/Fには、XMLを積極的に利用した。ただし、既存のリレーショナルデータベースをそのまま利用できる。

Javaは、OS/CPUに依存しないことを言語の基本姿勢としている。したがって、開発環境の移動が容易で、移植性が高い。また、Javaによる開発は、J2EE(Java 2 Enterprise Edition)を核に、Webアプリケーションの標準となりつつある。さらに、J2EEに利用できる多くのフレームワークや部品がフリーで公開されている。これらを再利用することで、開発のコストを削ることができる。

このシステムは、今年度中に、基盤システムとAstro-E2向けの部分を完成し、Astro-E2のデータ公開で実運用に入る予定である。図2に検索の入力と結果表示画面の例を示す。

4 今後の開発項目

Astro-F 衛星向けの開発 2005年度の打ち上げ予定の赤外線天文衛星。かつてない感度でのサーベイをおこない、 $10^6 - 10^7$ 個の赤外線天体のカタログをうみだす。DARTSでは、この大規模カタログデータベースの作成と検索システムの開発、データ配布をサポートする。打ち上げまでに、天文データに適した大容量(> 10 GBytes)データベースを管理するための基盤技術を開発する必要がある。

データ解析センター機能の拡充 ここ数年、天文学などの宇宙科学を専門とする国内の研究室の数が増えている。また、データ解析を必要とするのは、必ずしもその分野の専門家だけではない。さらに天文の分野では、多波長の観測を合わせることで、新しい発見がありえる。すなわち、宇宙科学データのユーザの数だけでなく、その多様性も増えている。しかしながら、それぞれの衛星データの解析環境は、大規模化・複雑化している。したがって、個々のユーザが少人数むけに解析環境を構築・管理するのは効率が悪い。このような手間を省き、宇宙科学データからのアウトプットを最大化するために、DARTSのデータ解析センター機能の拡充する必要がある。現在は、特にAstro-E2衛星むけおよび太陽観測の総合解析環境(SODA; 国立天文台などとの共同研究)²の構築に力を注いでいる。

“仮想天文台”への協力・対応 増大する天文データとあらゆるデータを容易かつ柔軟に利用したい研究者の要求に答えるため、仮想天文台(VO; Virtual Observatory)という計画が世界中で進められている。例えば、世界中の天文データを一元的に検索し、統一的に解析できるシステムを目指している。国内では、国立天文台のJVOグループ³がこの計画を進めている。我々は、数年前より国立天文台のデータセンター⁴との連携を進めてきた。実際に、多波長天文画像サービス(jMAISON)を共同開発している。また、スーパーSINET専用線を用いたデータセンター機能の連携・結合を進めている。以上に加え、宇宙研の天文データをVOシステムに提供する準備をJVOの協力のもとに進めている。特にAstro-F衛星のデータ公開時には、VOの利用は必須である。

太陽・地球物理系データ向けの開発 他機関と共同で、より柔軟性・拡張性の高いシステムの開発を計画している。Solar-B衛星向けのシステム開発も主要な課題である。

²www.cc.nao.ac.jp/soda/

³jvo.nao.ac.jp

⁴dbc.nao.ac.jp