

「すぎく」データ処理における国際協力

尾崎 正伸

〈宇宙航空研究開発機構 宇宙科学研究所 〒252-5210 相模原市中央区由野台3-1-1〉
e-mail: ozaki.masanobu@jaxa.jp



「すぎく」衛星から得られたデータの処理・配布は、「あすか」の経験と反省に基づき日米で共同かつ並行作業として定義され、実施されている。この経緯と採用した規則・実際の処理の流れについて紹介する。また、このシステムを開発する際に生じたいくつかのトラブルと、その反省に立った次に注意すべきことについても触れる。

1. はじめに

科学から離れて運営という観点から眺めたとき、「すぎく」(ASTRO-E2)プロジェクトの特徴の一つとして実務の国際分担による負荷分散が挙げられる。これは衛星を作ることにとどまらず、観測立案やユーザーサポートにまで及ぶ。その最も緊密かつ大規模な(と自負している)例として、筆者が担当してきた配布データ生成と解析ソフトウェアの開発・維持に関して紹介する。「すぎく」は世界中からの観測提案を受け付けデータを配布しているが、実際の(広義の)運用は日本と米国で行っている。したがってすべては「日米」二国間での話なのだが、内容は米国相手固有のことではないのでこのような表題とさせていただいた。

2. 「あすか」の経験

「すぎく」を語るうえでは、先代の「あすか」の経験に触れないわけにはいかない。「あすか」も日米協力で開発と運用が行われ、衛星の構成のみならず運営形態としても「すぎく」の先祖と位置づけることができる。

「あすか」のデータは衛星から降りてきた情報をほぼそのまま配布する形を取り、解析する際は

それを解読して科学データを取り出しFITS形式にするところからユーザーの手に委ねられていた。このデータ取り出しツールは衛星打上げに先立って作られ、解析ツールの一つとして提供されていた。そのほかのいろいろな解析ツールは、当初は米国や欧州の既存ツールを流用しつつ必要なものを日米の研究者が草の根的に補い、最後には現在のHEAsoftに一式そろえられたと記憶している。ツールは最初からインターネットで配布され、データは最初は衛星フォーマットをテープ配布だったが最後にはFITS形式にされたものがさらに定型的な解析まで施されてオンラインでNASAサイトから手に入るようになった。また、NASAは後期には頼まれずとも世界中の全観測者にCD-ROMでデータを配布した。

このような経緯をたどる中で、プロジェクトに最初から携わっていたわけではない世界の(特に米国の)一般の観測者の中には、「あすか」が米国主導のプロジェクトであるかのような誤解も生じたい。これは、NASAのASCA GOF (Guest Observer Facility) が積極的にソフトウェアの開発とデータの配布・ユーザーサポートを行ったからだと思われる。また、米国から配布されたデータがFITS形式のみを出発点としているせいで、日本の研究者が衛星フォーマットから直接行った

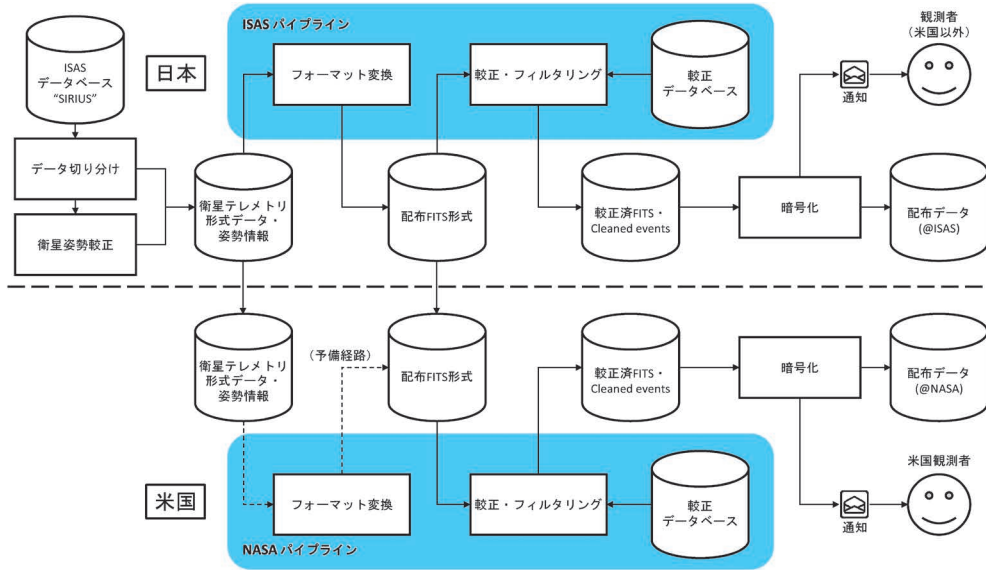


図1 「すざく」配布データの処理の流れ。日本で観測ごとに切り分けて配布形式のFITSに変換してからは、日米で全く同じ処理を施し同じデータをもつ。万が一の事故に備え米国にも衛星テレメトリ形式のデータを保持するが、通常の処理には使用しない。

解析^{*1}に対して、FITS形式では得られないインサイダー情報（そのようなものはそもそもなかったのだが）を使って研究を行ったのではないかと痛くもない腹を探られるような事態も生じた。

3. 「すざく」の戦略

そこで「すざく」のデータ配布および解析環境を整えるときには、このような問題が生じないようにしようというのが目標の一つになった。結果として作られた処理の流れを図1に示す。

これを決定するときに設けた規則は、

- テレメトリフォーマットは非公開とし、テレメトリデータそのものは配布しない。チーム関係者に対しても科学目的では非公開とする。すべての解析は日米で協議して正式に定めたFITS形式の標準フォーマットに変換したデータから始める。テレメトリデータか

ら標準フォーマットへの変換は宇宙研の専用工程でのみ行う。

- 「あすか」で有効だった、品質の良いデータのみを抜き出したサブセット（cleaned eventと呼ぶ）も全データとともに配布する。これにより検出器チームのデータ選別ノウハウを全ユーザが等しく使えるようにする。
- 標準フォーマットから先の配布のためのデータ処理は日米で並行して同期を取って行う。配布（観測提案者への案内）も米国向けは米国から、日本および他の全世界向けは日本から行う^{*2}。
- 解析ツールはすべてNASA HEASARCからHEASoftの一部としてリリースする。中身はエキスパートである検出器チームが供給する。というものであった。これらの理念を実現するため、日米でデータ配布業務の担当を決め、その間

^{*1} 当時の計算機環境では、たくさんのデータを解析しようとするときに途中フォーマットを経ないことは、ソフトウェア環境的・処理時間的・ディスクリソース的に大きなメリットがあった。

^{*2} データそのものは日米どちらのサイトからでも同じように取得できる。

で細かいファイル名やフォーマット、データのハンドリング方法やデータ公開タイミング、バージョン管理規則、ソフトウェアのデバッグ方法など細々としたことまで衛星打上げ準備と並行して設定し、各検出器チームにソフト開発を依頼した。これに加え、衛星からのデータを日米並行処理部へ供給するまでの部分は宇宙研で開発することになった。

4. デスマーチ

データ取り扱いの大方針はこのように打上げ前に定まったが、実装に落としていく段階がたいへんだった。時間も人手もないことから、「あすか」データ処理でNASAで使われていたソフトを極力流用する方針を取ったが、その関係で難航したのがデータに種々の処理を施し配布できる段階にまでもっていく「パイプライン」と呼ばれる部分だった。動いているものを使うのが早いだろうというNASA側の強い意向もありこのようになったのだが、実はこのソフトを書いた人物はすでにNASAを去っていた。悪いことに、ドキュメントはその人物の哲学により最初から全く存在せず(情報はすべてコメントの形でコードに埋め込み、という哲学だと聞いた記憶がある)、あるのはNASA環境で動いていたそのままの設定のソフト一式とASTRO-E打上げ前に行われたインタビューのメモだけという状態だった。またソフト自体も当時としては極めて趣味性の強い特殊な構造だった。ソフトの引き継ぎを行った経験がある方はおわかりだろうが、引き継ぐためには根底に流れている「思想」をまずしっかり理解しなければならず、それなしで集めた情報は思想を理解するまではほとんど役に立たない。思想がわかれば書いてあることも理解できるのだが、その中身はすでにこちらがもっている情報になっているので実質的な役には立たない。状況は米国側でも同じ

だったようで、米国側が面倒をみることになってきたこのソフトのセットアップは結局日米独立に行うことになり、日本では起動方法や環境設定から試行錯誤する事態となった。

反対に、過去の資産の流用を拒否してうまくいったのが解析ツールの実装である。NASA側はHEAsoftのソフトウェアテンプレートを使うことを強く要求してきたが、日本側はすでに使っていた検出器開発用ライブラリ^{*3}との互換性を盾にこれの拒否を貫き、ユーザから見えるインターフェイスの互換性だけを保つようにした。この結果、短時間でのリリース・デバッグと保守期間に入ってから各検出器チーム内での担当者の世代交代が円滑に行われたように思う。

当時のX線天文学コミュニティは、詳細な仕様書を作って外部に仕事を委託するという開発は極めて稀だったように思う。自分達で必要なものを作れる腕があったということだが、実のところ他人に適切に仕事をしてもらうための仕様の書き方を知らなかったというほうがより正確だろう。加えて、日本ではソフトウェアはハードウェアのおまけという雰囲気はまだ引きずっていた時代なので、ソフトウェア専門部隊がなかった。したがって、ソフトを書く能力のある人が他の仕事に加えて手掛けることになり、打上げ後のデータが出始める時期に駕籠舁き草鞋編みとして追加の仕事を引き受けることになった。

「すぎく」やその前身のASTRO-Eが作られた2000-2005年頃は、地域的に分散した集団で大規模なソフトウェアを素早く開発する手法が世界を見回してもようやく出そろった頃で、したがってソフトウェア開発が本業ではないわれわれの開発は手法的にいろいろと泥臭く時間のかかるものであった。その一方で、出てきたデータは早急に「すぎく」チームで評価して半年後の公募フェーズ開始までに一通りの較正を完了させなければなら

^{*3} 初期にはFORTRANで、後にはCで実装されていたが、オブジェクト指向の考え方を前面に押し出したインターフェイスで作られていた。

らなかった。したがって、解析ソフトは一般公開品質に達する前に内輪で使われ、パイプライン構築担当はデータ解析をする時間が全く取れないという状況に陥った。限られた数のソフト開発者がデータを解析する人々から早くツールを出せとせつつかれる、非常に偏った状態であった。この反省から、後にASTRO-Hではソフトウェアチームが検出器チームと完全に並立した部隊として定義され、打上げの何年も前からそれだけに集中して開発を行う体制となった。

ともかく、そのような状況なのでパイプラインを使った正規ルートでのデータ配布がまともに走り出したのは打上げから10カ月以上経ってからだった。走り出したとはいってもその先もさまざまなソフトの改修や追加があり、それに対応すべく各ソフトやパイプラインの担当者は相変わらずほとんどデータに触れない日々が続いたのだが、ともかく対外公約であるAO開始とそれに伴うデータ配布は（当時の水準に照らせば）ひどく責められなければならないほどには遅れず何とか始められたと思っている。さまざまな改修・追加が落ち着いたのは、打上げから実に2年が過ぎた頃だった。そして、その頃には次のASTRO-Hに向けた開発が本格化しており、ソフト担当者たちがデータに意識を振り向ける時間はやはりほとんど取れなくなっていたように思う。

日米で並行してパイプラインを走らせるという大方針は、ほかにもさまざまな苦勞を生じた。例えばソフト修正一つをとってみても、該当箇所を修正すればおしまいというのではなく、おかしなミスがないかを検証する仕組み^{*4}もNASA側の要請により途中から導入した。修正を反映させるときには日米のパイプラインを同期して停め、新ソフトで試験データを処理した結果を比較して違いが生じていないことまで確認してから再開さ

せる。このような作業を間違いなく手早く完了するための手順も自分たちで編出し蓄積することになった。ちなみに今に至るまで日本側担当者の交代は生じていないが、NASA側の担当者は幾度となく交代し、結果的にパイプラインはNASA製なのにすべてをきちんと知っているのは日本側のみ、などという時期もあった。先方の担当が急逝したときは、研究会会場（奇しくも「すざく」の国際研究会だった）から電話でNASA新担当者に知識を伝授するようなこともあった。

5. 戦い終わって

ともあれ、現在では「すざく」の配布プロセスとソフトウェアは安定し、衛星データそのものに問題があるケース以外は観測から数週間でデータを配布している。その間には使用する計算機の更新も複数回経験し、そのたびにパイプラインの外側の環境定義に少なからぬ労力を要求されつつも、対米インターフェイスには影響を出さず、データ配布への影響も日常レベルの遅延で乗り切ることができた。

最後に、このプロジェクトで得た教訓をまとめておく。

- 過去のソフトの流用を検討するときは、かならずコードそのものを精査してから可否判断すべし。特に、コードを読んでいない人の意見を鵜呑みにしてはいけない。
- 要件のアウトラインはあらかじめ文書化して、常に参照できるようにすべし。アウトラインのみならず、可能なら詳細仕様に至るまで文書化すべし。文書に詳しく定義されるほど、実装を外部に頼みやすくなり、研究者をこのような仕事から解放できる。
- 検出器と同じ体制でソフトウェアチームも定義すべし。メンバーは原則としてソフトウェ

^{*4} ユニットテストという。これはいったん安定して動き出せばソフトの開発や改修が極めて着実になるが、ユニットテストを考慮せずに作られたソフトに後から導入するときはテストのための入力や正解を正しく用意するのがとてもたいへんな作業になる。

ア専業とすべし。

謝 辞

「すざく」のデータ処理は、たくさんの人々の並々ならぬ献身と努力により作られ維持されてきた。その中でも特筆すべきなのは、それぞれの検出器の製作とキャリブレーションでたいへん大きな役割を果たされながらソフトウェア開発と維持にも継続的にご尽力くださった首都大学東京の石崎欣尚さんと埼玉大学の寺田幸功さん、データ処理と配布の日本側実作業をその一身に引き受けてくださった宇宙研の山岸泉さんである。この場を借りて特に御礼申し上げる。

International Collaboration in *Suzaku* Data Processing and Distribution

Masanobu OZAKI

*Institute of Space and Astronautical Science
(ISAS), Japan Aerospace Exploration Agency
(JAXA), 3-1-1 Yoshinodai, Chuou-ku,
Sagamihara 252-5210, Japan*

Abstract: *Suzaku* data processing and distribution scheme was defined based on the lessons learned of the ASCA project. The key rules of the scheme and resulting processing flow are described. Some trouble cases in the development phase and suggestions to avoid them are also given.