

第1回 SPARC Japan セミナー2020

「研究データ公開:フルオープンと制限公開の境界線」

宇宙科学研究所における データ公開・制限公開に関する議論の紹介

海老沢 研

(宇宙航空研究開発機構 宇宙科学研究所)

講演要旨



数年前まで、宇宙科学研究所 (ISAS) では、何を「データ」と呼ぶのか、その「データ」を公開してよいのかどうか、はっきりしないような状態が続いていた。2017年度より所内に「科学データ専門委員会」が設立され、正式なデータポリシーを策定した。そこでは対象となる「データ」を定義するとともに、原則としてすべてのデータは公開とし、非公開とせざるを得ない理由を限定した。また、非公開データについては、担当者と非公開期限を定めることにした。

データポリシーが策定されたおかげで、データに関わる議論は進めやすくなったが、具体的に、各データのどの部分を公開するか非公開とするか、非公開の場合はその期限をどれだけにするか、など判断が難しい場合も多い。それらについて、委員会で議論を進めながら、個別に判断している。

本講演では、そのような議論を紹介するとともに、同じ課題を抱えている他機関との情報交換も期待している。

海老沢 研

1986年京都大学理学部卒業、1991年東京大学大学院理学系研究科修了。宇宙科学研究所にて日本学術振興会特別研究員を経て、1992年よりNASAゴダード宇宙飛行センターにてX線天文学の研究とX線衛星データアーカイブ開発に従事。2001年から2004年までNASAから派遣され、ジュネーブ天文台インテグラル衛星データセンターに勤務。2005年より宇宙科学研究所教授。X線天文学の研究と大学院教育（東京大学天文学教室）に従事しつつ、科学衛星運用・データアーカイブユニット (C-SODA) にて科学データアーカイブDARTSの開発・運用を行う。宇宙科学研究所科学データ専門委員会委員長。



自己紹介

私が所属する宇宙航空研究開発機構 (JAXA) の宇宙科学研究所 (ISAS) は、1964年設立の古い研究所です。私は科学衛星運用・データ利用ユニット (C-SODA) に属しており、衛星データ処理、データアーカイブ基盤の開発・運用に携わっています。それから、科学データ専門委員会の委員長も務めています。私自身はX線天文学を専攻しており、2005年にJAXAに入所する前はアメリカ航空宇宙局 (NASA) と欧州宇宙機関 (ESA) でX線天文学の研究と衛星アーカイブ開発に従事していました。本日は専門の研究ではなく、

業務として行っているデータ整備・データ公開の話させていただきます。

ISASにおけるデータ整備・公開の状況

ISASには、宇宙理工学委員会という諮問委員会があります。これが所長に対していろいろな提言をします。2016年度に「宇宙科学研究所が保有するデータの取り扱いに関する提言」を提出しました (図1)。それ以前からいろいろな活動をしていたのですが、このあたりから物事が整理されていったので、そこから話を始めたいと思います。その資料は宇宙理工学メン

バーに限定公開していますが、そのエッセンスを紹介
します。これらの提言を実行に移す形で、ISAS を中
心として宇宙科学データの整備・公開が進められてい
ます。

まず、データの公開・非公開についてです。データ
の種類によらず汎用かつ有用なデータは公開し、長期
的に保管すること。非公開データは保管期限を定め、
期限後に扱いを見直すこと。データに関するさまざま
な事項を審議し、所に進言するための有識者委員会を
設けること。また、無条件に商業利用や宣伝利用して
も良い科学データを置く公開領域を設けてはどうかと
いった提言をしています。この提言に従って、2017
年に ISAS 内に科学データ専門委員会が作られ、私が
今まで委員長を務めてきました。

「宇宙科学研究所のデータポリシー」の紹介

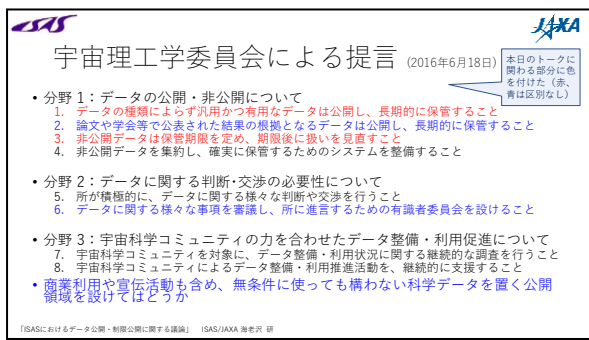
この科学データ専門委員会が、「宇宙科学研究所の
データポリシー」を策定しました(図 2)。諮問委員
会が所に対して行った提言に留まらず、所としてフォ
ーマルなデータポリシーを作ろうということで、議論

を重ねました。日本語・英語で ISAS のホームページ
より公開しています。ただ、ホームページで公開する
だけでは正式な文書になりませんので、文書番号を付
けて JAXA 内で公式な社内文書にしています。このよ
うに、フォーマルなデータポリシーを定めました。そ
のうち、本日の話に関係する部分を紹介させていただきます。

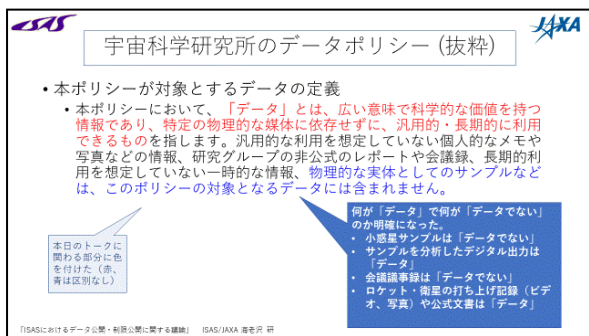
まず、データについて定義しています。『データ』
とは、広い意味で科学的な価値を持つ情報であり、特
定の物理的な媒体に依存せずに、汎用的・長期的に利
用できるものを指します、「物理的な実体としてのサ
ンプルなどは、このポリシーの対象となるデータには
含まれていません」としています。例えば私たちの研
究所で、「はやぶさ」がイトカワのサンプルを採って
きました。そのサンプルはここではデータとはいいま
せん。しかし、そのサンプルを分析したデジタル出力
はデータです。また、いろいろな会議の議事録も保管
されているのですが、それはデータとは呼びません。
その一方で、ロケットや衛星の打ち上げ記録などの公
式文書はデータということになります。

このように、まずは、何がデータで何がデータでな
いかをはっきりさせました。以前はそれがはっきりし
ていなかったもので、そもそも何がデータで何がデータ
ではないかという議論もできませんでしたが、それが
はっきりしてきたということです。

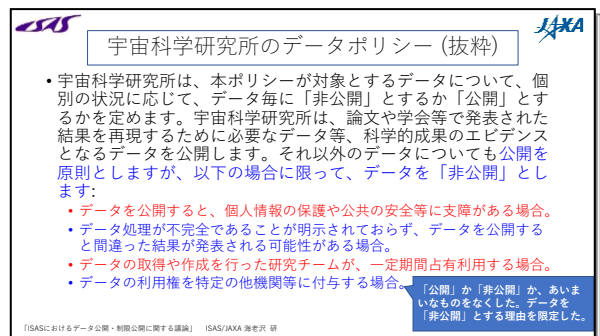
データポリシーとしては、まず公開を原則とします
(図 3)。しかし、次の場合に限ってデータを非公開
とします。データを公開すると個人情報の保護や公共
の安全等に支障が出る場合。データ処理が不完全であ



(図 1)



(図 2)



(図 3)

ることが明示されておらず、データを公開すると間違
った結果が発表される可能性がある場合。データの取
得や作成を行った研究チームが、一定期間、占有利用
する場合。データの利用権を特定の他機関等に付与す
る場合です。このように、公開か非公開かが曖昧なも
のをなくし、データを非公開とする理由を限定しまし
た。

非公開データについては、利用範囲と非公開期限を
定め、その後、公開に移行するか、非公開を続けるか、
あるいは廃棄するかを判断します。これは非常に大事
で、後々、いろいろな場面で効いてきます(図4)。

非公開といっても、誰も使わない非公開はあり得ま
せん。また、永久に非公開のデータも原理的にあり得
ませんので、どこかで廃棄する、あるいは公開に移行
することになります。非公開を長く続けるという手も
ありますが、極端な話、100年間非公開というのは考
えにくいです。私は天文学者ですが、新しい星が発見
されたとき、100年前の写真を見て、100年前の空に、
実はその星が受かっていた、ということはよくあるこ
とです。ですから、100年後に宇宙科学研究所のデー
タを使うことは十分あると考えています。

公開データのポリシーは図5のようになっています。
これは後で気が付いたのですが、世界的に公開データ
を Findable, Accessible, Interoperable, Reusable にしよう
と言う FAIR principles というものがあるそうです。私た
ちが自分たちの頭で考えたものが、こういうものと合
致しています。

公開データ利用の際のルールは図6のとおりです。
公開データは、原則として、営利目的、非営利目的を

問わず、複製、送信、加工も含め無償で利用できる。
また、必ず出典を明示すること。宇宙科学研究所は、
利用者がデータを用いて行う一切の行為について何ら
責任を負うものではない。このように非常に寛大なル
ールで、政府標準利用規約(第2.0版)やCC BY 4.0
など実績のある外部のルールを引用しています。公開
データのルールを定めることにより、データ利用が進
み、安心してデータを公開することができます。私た
ちは ISAS のデータポリシーに則っている、このデー
タポリシーは JAXA の正式文書としてフォーマルなも
のであるということで、物事が着々と進んでいます。

公開データ・データ利用の例

ISAS のデータポリシーに沿って、着実にデータ公
開が進んでいます。例えば「かぐや (SELENE)」は
月を周回する衛星です。「かぐや」については、NHK
が作ってきたきれいに整備されたハイビジョンカメラのデー
タがあったのですが、それが公開か非公開かが長い
間、曖昧な状態でした。それも、データポリシーに従
って、公開するか、非公開だったら何年間非公開かと

(図5)

(図4)

(図6)

いうことを決めなければいけません。そこで所長の命で、JAXA の広報部の方なども含めて NHK と交渉し、公開することになりました。公開か非公開か曖昧なものを作ってはいけない、というポリシーが効いた例です。

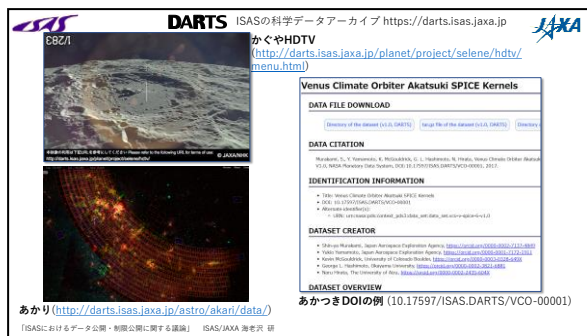
また、「あかり」衛星の場合には、未処理の生データがたくさんありました。このデータは貴重なものなので公開しよう、そのためにはデータポリシーに従って使える状態にして公開しなければいけないということで、予算を付けて、高次処理をして公開しました。

それから、データへの DOI 付与も進んでいます。

現在稼働中の「あらせ」衛星のデータ処理は名古屋大学が主に行っていますが、過去の ISAS の衛星データがいろいろな大学に散らばっています。これについてもデータ整備・公開を進めなければいけないということで、ISAS が予算を付けて支援しています。このようにしてデータ整備・公開が着々と進んできています。

いくつか例を示します(図7)。「かぐや」のハイビジョンのデータです。非常に美しい映像で、JAXA と NHK のロゴが入っています。公開できなかったら本当にもったいないことでしたが、きちんと公開できています。「あかり」のデータも素晴らしいもので、きちんと整備して公開しました。それから、DOI が付いた例としては金星を回る「あかつき」のデータがあります。DOI をたたくと「あかつき」のデータが出てきます。これらのデータもよく利用されています。

また、データの商業利用も進んでいます(図8)。これはグリー株式会社と一緒にやった仕事です。やは



(図7)

りプロのプログラマーはすごいなと思ったのは、データの使い方を知らせると、すぐに VR のコンテンツにするのです。宇宙を題材にした子ども向け VR 体感サイエンスツアー「ありえな LAB」というものです。もちろんデータ利用にお金は取りませんが、必ず私たち JAXA、ISAS の DARTS (データベース) に acknowledge していただき、CC BY 4.0 の規定に従って使用していただいています。

一般利用の例としては、最近気付いたのですが、YouTube にきれいな「あかつき」のデータが加工されてアップされています(図9)。かなりよくできるアマチュアの方のようで、「DARTS は JAXA のいろいろな探査機の資料を見ることのできる宝箱でした」と書いていただいて、とても嬉しかったです。一般の方も ISAS のデータを使って、何か科学的発見をされるかもしれません。公開して使ってもらわなければ何が出てくるか分からないということで、データを公開しています。



(図8)



(図9)

「制限公開」に関わる議論の紹介

ここからが本題になるかと思いますが。制限公開に関わる議論は非常に難しく、私たちも悩んでいることがありますので、本日の研究会でいろいろな専門家の意見を聞きたいと思っています。

まず私たちのデータポリシーですが、公共の安全等に支障がある情報は非公開にします（図 10）。例えば周波数情報は非公開です。人工衛星と地上で通信しますが、その周波数で何かを送って人工衛星を止めるなど、原理的にはできるかもしれません。それから、人工衛星と地上との間の通信を邪魔することができるかもしれません。過去の衛星なら問題ないだろうと思っても、同じ周波数を現在の衛星にも利用していることがあるそうです。工学の先生方の話を聞くと、やはり周波数は出したくない、出す必要もないということでした。

JAXA リポジトリが国立情報学研究所（NII）にあります。ISAS 報告で、M-V型ロケットについての科学論文があります。これは公開されていますが、「〇〇のレーダ・トランスポンダ」と白塗りになっています。隠されている部分は周波数です。周波数情報は非公開という具体的な例で、その判断にはあまり困っていません。

次が、困っている例です（図 11）。テレメトリデータと工学データです。1970年代からのテレメトリデータと、それを分かりやすくした、探査機の位置・姿勢制御、温度、電圧などの工学データが保管されています。これらをどうするべきか。さすがに直ちに破棄はないでしょう。しかし、先ほどのロジックで永久非

公開もないでしょう。ただ、公開すると、公共の安全や研究チームの一定期間占有利用に抵触しないか。

「はやぶさ」で有名な先生にも意見を伺ったのですが、そういうものは公開しなくていいのではないかという意見でした。

「はやぶさ」の地球再突入直前には、機微なオペレーションがあったそうです。そこでどんなことをしたかを人に知らせると、衛星や探査機メーカーのノウハウも知られてしまいます。そこで非公開にしようとなったのですが、永久非公開はあり得ない。では何年待って公開するか。これはまだ決まっていません。私たちの委員会でも散々議論しているのですが、困っているところです。

もう一つ困っているのが、「宇宙科学資料データベース」です（図 12）。ISAS とその前身の研究所の1950年代からの数十万枚の画像、映像が保管されています。かつては ISAS に映像記録班が存在していました。昔は国家公務員として写真を撮ったり、映像を撮ったりする専門的な方がいて、プロフェッショナルなクオリティで、ほぼ全てのロケット・衛星開発記録

テレメトリデータ・工学データ

- 1970年代からの科学衛星の「テレメトリデータ」（生データ）が保管されている
- テレメトリデータを解読した、探査機の位置・姿勢制御、温度、電圧などの「工学データ」が存在する。
- 終了した衛星のテレメトリデータ・工学データをどうするか？
 - 破棄はないだろう
 - 永久非公開はありえない
 - 公開すると「公共の安全」や「研究チームの一定期間占有利用」に抵触しないか？
 - たとえば、「はやぶさ」の地球再突入直前のオペレーションがわかってしまう
 - 衛星や探査機メーカーのノウハウも知られてしまう
 - 一定期間は非公開としたら、その期間は？(10年～30年?)

ISASにおけるデータ公開・制限公開に関する議論 | ISAS/JAXA 海老沢 研

(図 11)

「公共の安全等に支障がある」情報

- 「周波数情報」は非公開
- <http://id.nii.ac.jp/1696/000333/61/> など
- 周波数情報を白塗りにして公開

ISASにおけるデータ公開・制限公開に関する議論 | ISAS/JAXA 海老沢 研

(図 10)

「宇宙科学資料データベース」

- ISASとその前身の研究所の1950年代からの数十万枚の画像、映像が保管されている
- かつてISASに「映像記録班」が存在した
 - プロフェッショナルなクオリティ
 - ほぼすべてのロケット・衛星開発記録がデジタル化されている
- 現在、JAXA内でアカウントを申請して閲覧・利用できる
 - Dspace というシステムに載せていて、検索も可能

ISASにおけるデータ公開・制限公開に関する議論 | ISAS/JAXA 海老沢 研

(図 12)

を取っていました。もちろん昔は写真、フィルムでしたが、膨大なお金をかけてこれを全てデジタル化しました。現在、宇宙科学資料データベースは JAXA 内でアカウントを申請して閲覧・利用できます。DSpace というシステムに載せているので検索も可能です。これは一部公開していますが、まだ全部は公開していません。

他にも非公開資料がたくさんあります。例えば 1950 年代のカップロケットや、日本初の人工衛星「おおすみ」打ち上げ後の街の様子など、写真が何十万枚もあるわけです。これらの画像や映像の公開には、著作権、肖像権、公共安全、メーカーのノウハウなどの観点から問題がないか、今、議論しているところです。

私たちの委員会では、宇宙科学資料データベースを 30 年後には公開することを決めました。世の中には 30 年ルールというものがあるそうで、いろいろな公文書が 30 年で公開されています。ただ、その多くはまだ公開できていないというのが現状です。担当の方に聞くと、事なかれ主義という言葉は悪いですが、絶対に問題を起こしたくないという意識が高いように思われます。一方で、数十年にわたり、多額の税金をかけて収集・整備した資料を公開しないことは許されるのかというと、やはり公開しなければいけないのだと思います。責任がある誰かに、びしっと「公開すべきだ」と言って貰えれば良いのですが。

今こういう問題があり、困っているということをご紹介しました。ぜひ多くの皆様のご意見を頂ければと思います。