

資料2

国際的動向を踏まえたオープン  
サイエンスの推進に関する検討会  
(第7回)

平成30年10月26日(金)

# 「宇宙科学研究所(ISAS)の データポリシー」策定の経緯 について

2018年10月25日

宇宙航空研究開発機構(JAXA)・宇宙科学研究所(ISAS)

宇宙物理学研究系

科学データ専門委員会委員長

海老沢 研

# 話の内容

1. ISASのデータポリシー策定に至る背景と周辺状況
2. ISASのデータポリシー策定において議論したこと

以下、（余裕があれば）議論の材料として：

3. データポリシーに関係するJAXA内の状況
4. データポリシーに関して国に期待すること

# 1. ISASのデータポリシー策定に至る背景と周辺状況

- 私の簡単な経歴
  - 1991年、理学博士(天文学)、専門は主に人工衛星を用いたX線天文学
  - 1992年からNASAにてX線天文データアーカイブ開発・運用
  - 2005年からISAS/JAXAにて宇宙科学データアーカイブ開発・運用
- 私が見てきた「オープンサイエンス」
  - インパクトの大きな間違った結果はオープンでないデータから生まれる
    - 宇宙背景放射の黒体放射からのずれを「発見」1988年
    - 超新星残骸からTeVガンマ線の「発見」1998年
  - 今では、**短い占有期間の後に天文衛星データは「オープン」**になる
    - データがオープンであることはあたりまえ。結果を再現または検証できる。
  - **データをオープンにすることで多くの成果が生まれる**
    - 日本の「あすか」衛星のアーカイブを日米協力で開発。30以上の国から1500本以上の投稿論文
  - NASAデータポリシーは非常にオープン(NASA Policy Derivative 2230.1)
    - NASAは著作権を主張しない ([http://jda.jaxa.jp/nasa\\_guideline.php](http://jda.jaxa.jp/nasa_guideline.php))
    - ハッブル宇宙望遠鏡の画像が宗教の宣伝パンフレットに！
  - JAXAのデータは(NASAに比べて)使いにくい
    - JAXAは著作権者
    - 原則として許可制、商業利用不可(<http://jda.jaxa.jp/service.php>)

# 1. ISASのデータポリシー策定に至る背景と周辺状況

- 2008年 ISAS内に「科学衛星・運用データ利用センター」設立
  - Center of Science-satellite Operation and Data Archive (C-SODA)
  - **C-SODA運営委員会**でデータに関する様々な議論、レポートを作成・公開(<http://c-soda.isas.jaxa.jp/bibliography/>)
    - 「科学データの公開・利用について」2011年
    - 「望ましい科学データの公開・利用のありかたについて」2013年
- 2014年、ISAS所長から宇宙理工学委員会に対して諮問；「ISASが保有するデータの取り扱いに関する提言のお願い」
  - **宇宙理工学委員会内**にタスクフォースを設立、議論
  - 2016年6月、宇宙理工学委員会から答申；「ISASが保有するデータの取り扱いに関する提言」
  - ISAS内に「データに関する様々な案件を審議する専門委員会」を設立すること
- 2017年、「**科学データ専門委員会**」でデータポリシーを議論
- 2018年3月、研究所会議で承認
- 2018年5月、日本語・英語でISAS HPの下で公開
  - <http://www.isas.jaxa.jp/researchers/data-policy/>
  - <http://www.isas.jaxa.jp/en/researchers/data-policy/>
  - 英文は日本人の英語専門家、ISASのイギリス人研究者、アメリカ人研究者がチェック

## 2. ISASのデータポリシー策定において議論したこと

- 以下、「ISASのデータポリシー」と対比させて、上記委員会で議論した内容を示す
- データポリシーは赤い四角で囲んだ

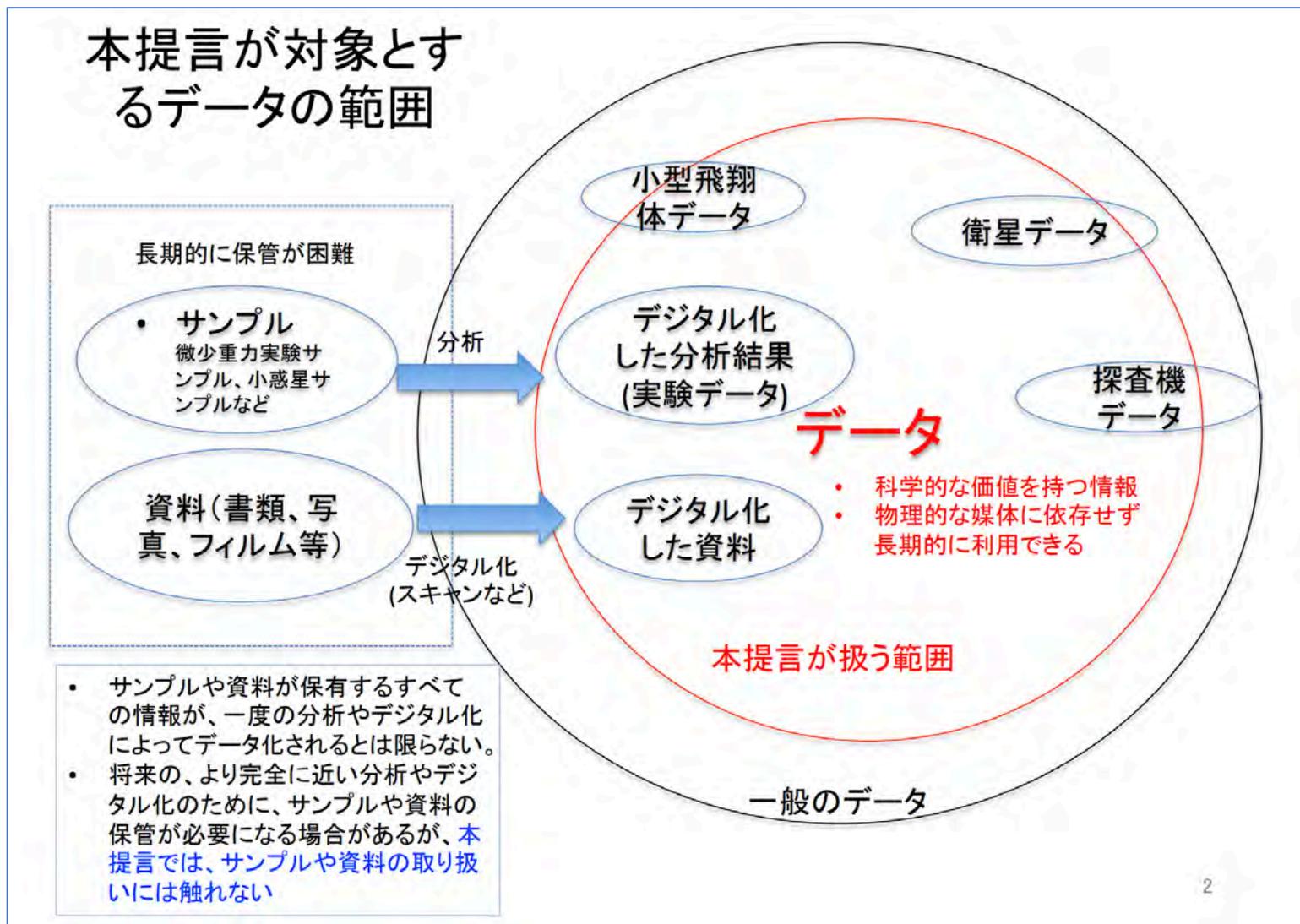
# 宇宙科学研究所のデータポリシー

(全文；脚註は除く)

## 1. 本ポリシーが対象とするデータの定義

以下では、「データ」とは、宇宙科学研究所が保有する、**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。**汎用的な利用を想定していない個人的なメモや写真などの情報、研究グループの非公式のレポートや会議録、長期的利用を想定していない一時的な情報、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ません。

# 何が「データ」で「データでないか」の判断基準を提示



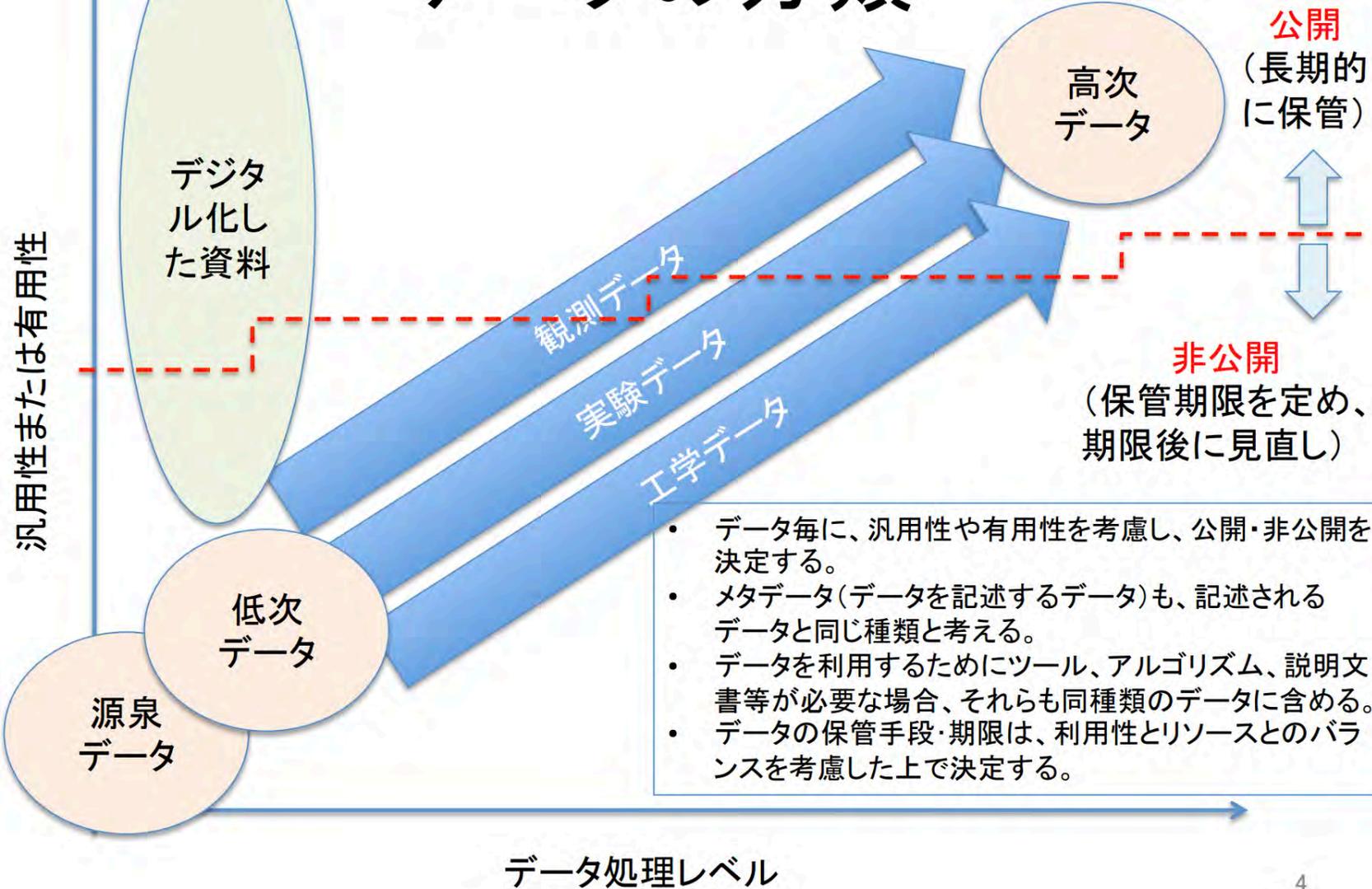
# 「何がデータで何がデータでないか」の判断は非常に重要

- データポリシーができるまでは、何がデータで何がデータでないかの判断ができなかった
- データポリシーに基づいて、科学データ専門委員会にて実のある議論ができるようになった
- 「宇宙科学資料データベース」に含まれる...
  - ロケット開発ミーティングの議事録は「データではない」
  - ISAS初期のロケット開発、打ち上げの記録写真は「データ」
  - ISASの研究者へのインタビュー映像は「データ」

以下に、ここで扱うデータの代表的な例と、簡単な説明を述べます：

- **源泉データ**：衛星テレメトリなど、定まったフォーマットで判読されることが必要で、データ処理の源泉となるもの。これから適切なデータ処理を行うことによつて、観測データや工学データなどを生成することができる。
- **観測データ**：衛星、探査機、大気球、観測ロケットなどによって、天体や宇宙現象など、制御できない対象の物理的状态を測定した数値データ。多くの宇宙現象は顕著な時間変動を示すので、同じデータを再現できないことがある。
- **工学データ**：衛星、探査機、装置の軌道、姿勢、温度など、データを取得する側の物理的状态を記述した数値データ。
- **実験データ**：観測者が、対象に何らかの意図的な操作を加えることによって測定した数値データ。多くの場合、実験を繰り返すことにより、同じデータを再現することができる。
- **シミュレーションデータ**：観測データ・工学データ・実験データなどを模擬するために、計算機によって生成されたデータ。計算を繰り返すことにより、同じデータを再現することができる。
- 探査機で得られた小惑星サンプルや微少重力実験で得られたサンプルの**分析結果を定量化した数値データ**や、探査機が測定した**天体の形状を再現した数値データ**など。分析・再現精度が向上することによって、改訂されることがある。
- **汎用的・長期的な利用を意図して作成された**、デジタル化された書類、写真、画像、映像など。

# データの分類



また、各データを記述するためのデータ（メタデータ）や、データを利用するために必要なツール、ソフトウェア、アルゴリズム、説明文書等も、対応するデータに準じて扱います。

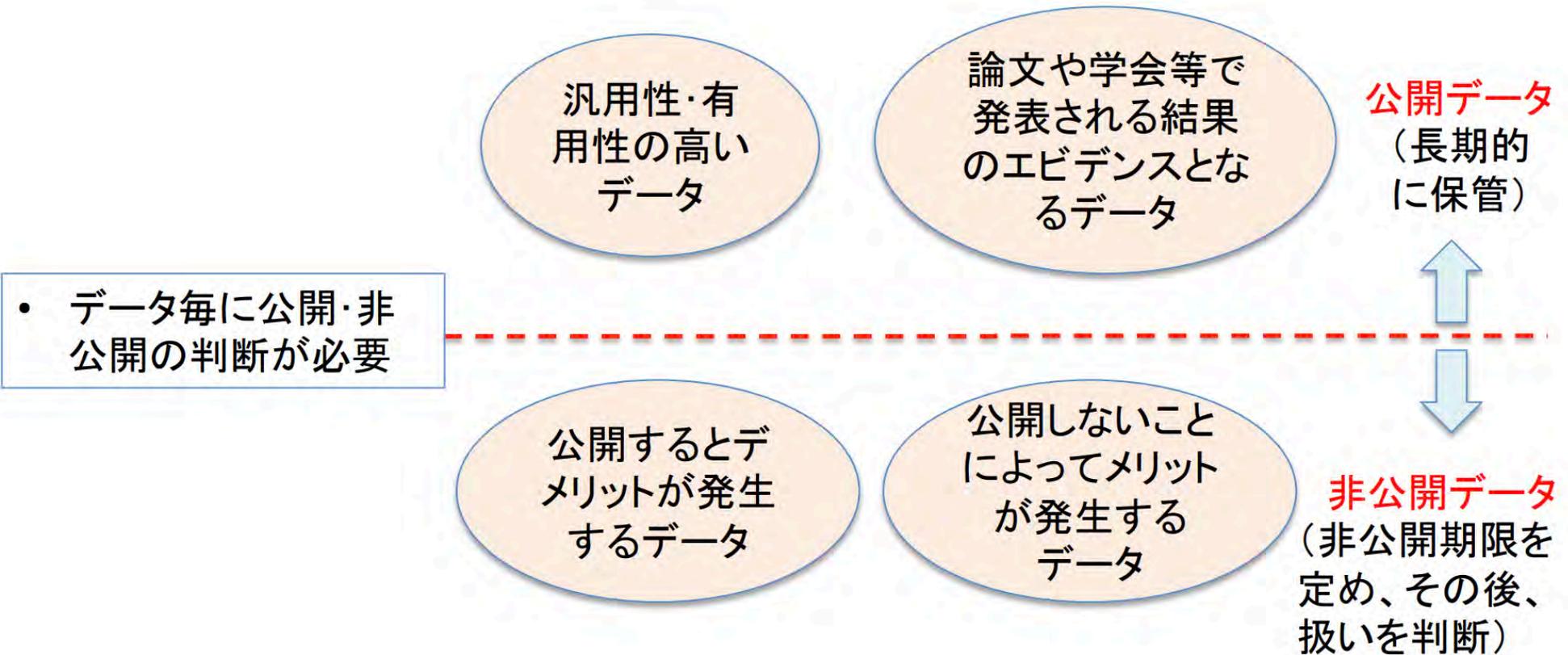
## 2. 非公開データ、公開データの定義

宇宙科学研究所は、本ポリシーが対象とするデータについて、個別の状況に応じて、**データ毎に「非公開」とするか「公開」とするかを定めます。**

- 非公開データ：限定的な範囲・期間における利用を意図して作成されたデータ。
- 公開データ：非公開データ以外の、誰でも自由に取得し、利用できるデータ。

# データの公開・非公開について

- 公開データの利用に際して、対価は要求しないが、出典(クレジット)の明示を要求する
- 出典をどこにするかは状況による(ISAS、JAXA、プロジェクト、DARTS、他機関など)



# 公開か非公開かの区別は非常に重要

- 今まで、公開しても良いかどうかわからないデータがたくさんあった
- 今後は、そのような曖昧な位置づけは許されない
- ISASのデータは、必ず公開か非公開かの、どちらかに分類する

### 3. 公開、非公開の考えかた

宇宙科学研究所は、論文や学会等で発表された結果を再現するために必要なデータ等、科学的成果のエビデンスとなるデータを公開します。それ以外のデータについても公開を原則としますが、以下の場合に限って、データを「非公開」とします：

- 公開することによって、個人情報情報の保護や公共の安全等に支障がある場合。
- 公開することによって、デメリットが発生する可能性がある場合。
  - たとえば、データ処理が不完全であることが明示されておらず、それを使って間違った結果が発表される可能性がある場合。
- 公開しないことによって、メリットが期待される場合。
  - たとえば、データの取得や作成を行った研究チームに、一定期間占有利用させることによって、優位性を付与する場合や、データの利用権を他機関との交渉材料にするなど、戦略的な利用価値がある場合。

# 非公開データの例

- ロケットと地上局間の通信周波数 (公共の安全等に支障)
  - 公開文書についても、一部を墨消し
- 公開しないことによって、メリットが期待される例
  - 探査機のクリティカル運用に関わる工学的データ、
  - 膨大な計算によって得られた、将来の月着陸検討に関わるデータ等
  - 海外の宇宙機関との交渉などに、戦略的に利用できる可能性がある

- 宇宙科学研究所は、原則として、非公開データの存在を、それを非公開とする理由とともに、明示します。ただし、データの存在を公開しないことにメリットがあると考えられる場合には、その存在を公開しない場合もあります。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。
- 衛星、探査機、大気球、観測ロケットなどによる観測データについて、装置チームが機器校正を行ったり観測提案者が占有利用したりするための非公開期間が必要な場合、その目安はデータ取得後から約1年です。

# 「非公開」データの利用・保管について

- 誰も使わない非公開データ、は論理的にありえない！
- 非公開データを永久保管は意味がない！
- いままで非公開で何十年も放置されてきたデータがあった
- **非公開データの担当者、利用範囲、保管期限を定義することをルール化した。**
- 非公開期限が過ぎたら、公開に移行する、または廃棄する
- 「宇宙科学資料データベース」に含まれる過去の文書資料、写真、ムービー等は30年たったら公開、と定めた。

「恒久的に」という意見もあったが、議論の末、現実的な期間とした（「ぎんが」衛星のデータは1987年～）

## 4. 公開データのポリシー

宇宙科学研究所は、公開データが広く利用されることで科学の進歩につながるとの信念に基づき、以下の方針を実施します：

- 公知の知識のみで公開データを利用できるように、適切なデータ処理やデータの説明を行います。
- 公開データは、利用できる状態で、**長期間（最低30年）** 保管します。
- 必要な公開データを簡単に見つけ、使いやすくするためのサービスを無償で提供します。
- 永続的な識別子を用いて、公開データを引用しやすくします。

科学データ公開のFAIR principleに準拠

Findable  
Accessible  
Inteorporative  
Reusable

<https://www.nature.com/articles/sdata201618>

## 5. 公開データ利用の際のルール

- 公開データを利用する際には、以下のルールに従ってください。なお、このルールは**政府標準利用規約（第2.0版）**に準じたもので、**クリエイティブ・コモンズ表示4.0 国際 (CC BY 4.0)**と互換性があります：
- 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ます。ただし、一部のデータについては、宇宙科学研究所以外の第三者が用途を制限している場合があります。
- データの利用に関しては、「**宇宙航空研究開発機構・宇宙科学研究所**」または「**ISAS/JAXA**」と**出典を明示**してください。さらに、データによっては、その取得・整備・公開等に関わった個人や組織が出典の明示を要求している場合もあるので、利用者の責任において、それに従ってください。
- 公開データを加工して利用する場合には、加工したという事実を明記するとともに、できる限り、どのような加工を行ったかを示してください。
- **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ありません。**

# 公開データの利用について

- **改変可**。データ解析で改変するのは当たり前。
- **商業利用、無条件で可**
  - Nature, Scienceなどの商業誌は科学データを使った出版で商売をしている！
  - 過去に、JAXAに申請して「かぐや」データを使ったソフトウェア会社は、データ利用料金はゼロ(!)だが、毎年JAXAに販売実績報告が義務づけられている
  - 最近、グリー社と「かぐや」データを使ったプロジェクトを開始した
- **データの利用条件はつけない**
  - 「科学」なので、データをどう使って、何を言っても良い
  - データを利用した結果（論文、発表）は、科学的・客観的な評価を受ける。
  - 「間違った」結果の責任は利用者にある（何が間違いかどうかも自明ではない）
- 政府標準利用規約（第2.0版）、クリエイティブ・コモンズ表示4.0 国際 (CC BY 4.0)と同様の考え
  - データポリシーは自分たちで考えたものだが、広く用いられている「共通言語」を引用した。

## 6. データ保存の考え方

宇宙科学研究所は、原則としてすべてのデータを可能な限り保存しますが、コストやリソースの観点から、データを廃棄せざるを得ない場合もあります。以下に、データ保存についての考え方を示します。

- 公開データは、長期間保管する。
- 原理的に再現できないデータは、長期間保管する。
- データ処理によって観測データや工学データなどを再現するための源泉データは、できるだけ再処理が可能な状態で、長期間保管する。
- 原理的に再現できるが、再現に大きなコストがかかるデータは、できるだけ廃棄しない。
- データ処理、実験、計算等によって比較的容易に再現できるデータは、廃棄する可能性がある。

重要性高



重要性低

宇宙科学データの多くは原理的に再現できない（e.g. 超新星爆発）という特徴がある  
（e.g. 素粒子実験データは再現できるのでアーカイブしない）