議論）将来に向けた光赤外天文観測データアーカイブの在り方への提言

2021.9.9 光赤天連シンポジウム総合討論の時間

古澤 久徳@国立天文台（天文データセンター・ハワイ観測所）
光赤天連データアーカイブWG
世話人：古澤久徳，諸隈智貴，表 泰秀，川端弘治，山村一誠

この時間について

提言書：
https://drive.google.com/drive/folders/11kbYfGbFciqLD0XengcNFHcqE8zeU91j?usp=sharing

意見収集フォーム：
https://forms.gle/VtnF9cuUFWnnq39W6

- 経緯説明
- 提案の骨子の説明
- 提言への意見交換、声明にむけて

1年前（2020.9光赤天連シンポにて）

重要な話

- 日本の可視赤外観測データアーカイブの維持・改善を目指し、運用に対して提案を作ろうとしています
- 2021年3月に光赤天連の声明とすることを目指しています
- そのためのワーキンググループへの参加者を募ります

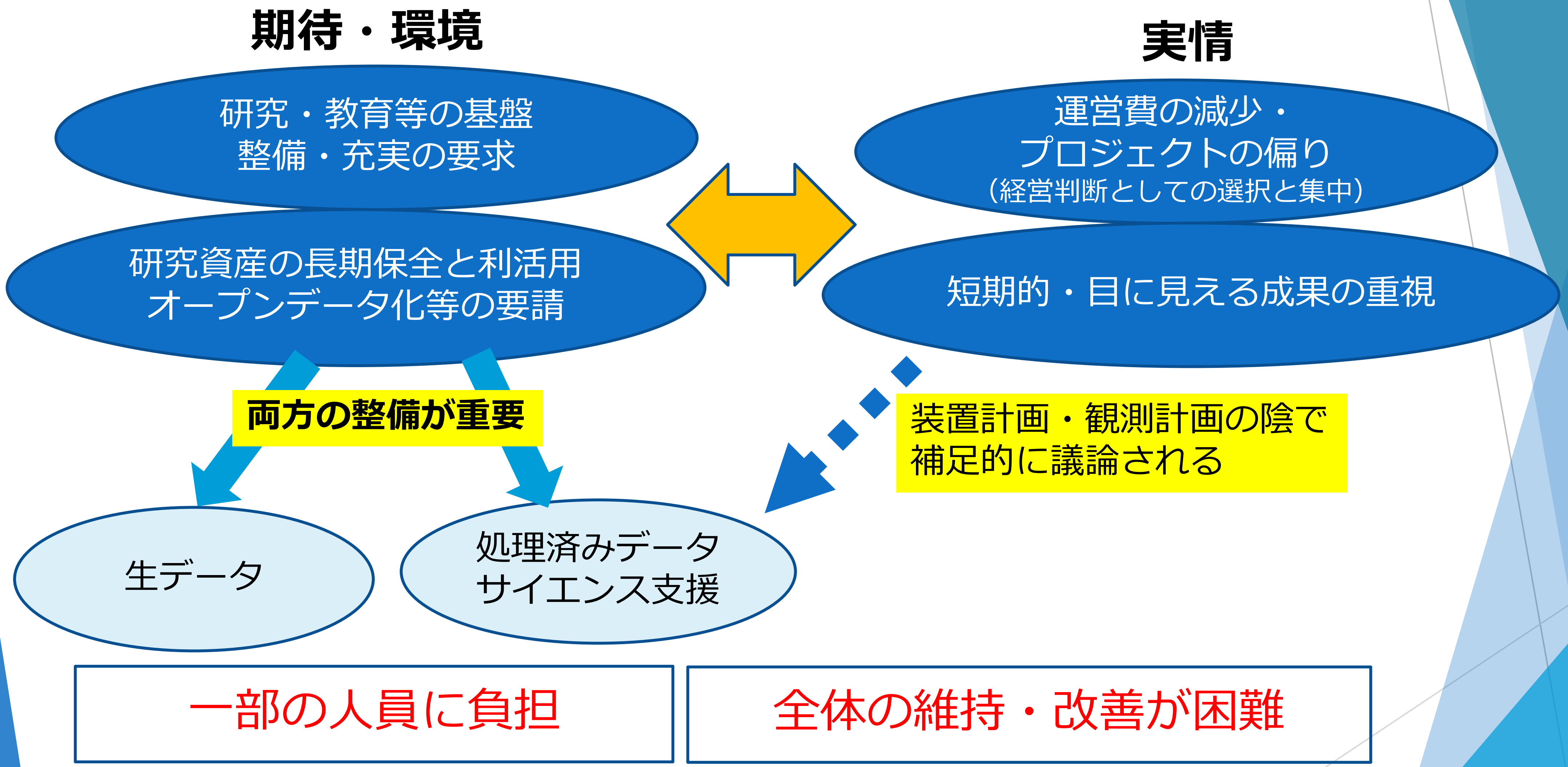
議論の時間にも活動に同意いただけるか意見分布を伺う予定です.

→ ご賛同をいただいて，活動を開始した.

データアーカイブ議論の経緯（１）
天文ビッグデータ時代とデータ運用

- HSC/PFS/Euclid/Roman/LSSTなど大サーベイ観測
- データをどう作り（生データ・処理）、どう保存し活用するのかが、成功の鍵であるとともに責務 → データアーカイブ

データアーカイブ議論の経緯（２）



- コミュニティでデータアーカイブ運用の意義や目標を共有できているのか？
- アーカイブ運用者含めてコミュニティが正しく努力出来ているのか？

データアーカイブ議論の経緯（３）

日本の天文学コミュニティ全体で考えるべき問題

→ 将来にわたる動機付けと目標設定が「今」必要！

人材の維持・育成の上でもクリティカル

- **データアーカイブの意義と目的の再共有**
  - 本当に必要だと思っているのか？
  - 何が必要なのか？　→ 皆さん必要だとは思っている！
- **それを実現するための手段と役割の明確化**
  - 誰がどこで何をすべきか　→ 各自の負担を増やすということではなく協力してメリットを最大にする

→ 自主的に，データアーカイブWGで議論し提言を作成

光赤天連声明として，天文コミュニティに対して
データアーカイブ運用の在り方を提言することを目指す

光赤天連データアーカイブWG

- 前身のワークショップ：2020年1月＆3月
- 活動期間：2020年9月から，これまでに9回の全体会合と多数の世話人・個別のグループでの議論を行った
  - フィードバック
    - 望遠鏡/装置チームからの意見収集，ハワイ観測所SAからとの議論，など
- メンバー（22名）
  - 古澤久德, Pyo Tae-Soo, 小杉城治，高田唯史，花山秀和，中島康，林将央，田中賢幸，岩田生，寺居剛，冨永望（国立天文台），大宮正士（ABC), 諸隈智貴，大澤亮，鮫島寛明（東京大），矢部清人（カブリIPMU），川端弘治（広島大），秋山正幸（東北大），野上大作（京都大），本田敏志（兵庫県立大），山村一誠，海老沢研（JAXA）（下線は世話人）

提言の主旨

- 提言先と目的
  - 光赤天連（我々自身）を含む**天文学研究コミュニティと関係機関**にむけて
  - 地上赤外天文**データアーカイブ（生・処理済）の運用の在り方**について提言を行う
  - これを光赤天連の声明とし，今後の具体的な実装議論へ向けた指針とする
- 背景
  - データアーカイブは観測天文学推進のための重要な研究基盤であり，光赤外天文学の将来計画や社会要請上もその必要性がますます高まっている
  - しかし，データアーカイブを維持・改善するためのコミュニティ内の協力体制・サポートや将来計画の枠組みでの議論が十分ではなかった
- 提言の骨子
  - データアーカイブに関わるコミュニティメンバーが，今後のデータアーカイブの**目指す姿と役割を共有**して運用計画に沿って協力することで，**これまで以上に有用なデータアーカイブ**を**安定的に運用**する
  - コミュニティや関係機関にはそのための**正当な評価と支援**をお願いしたい
  - この提言はデータアーカイブの在り方・**考え方の基盤を共有**することであり，この提言をもとに，**今後の具体的な実装を議論していく**

データアーカイブの意義と役割

- 科学的価値のある**データ資産の永続的な保管と有効な再利用による科学推進，社会要請への対応**
- 役割
  - <u>宇宙の記録</u>の保存と利用可能な状態での<u>公開</u>
  - <u>将来，未知の事象</u>の発見
  - 科学成果の<u>検証</u>
  - <u>科学・教育・社会活動の促進</u>

データアーカイブをちゃんと運用すると良いことがある

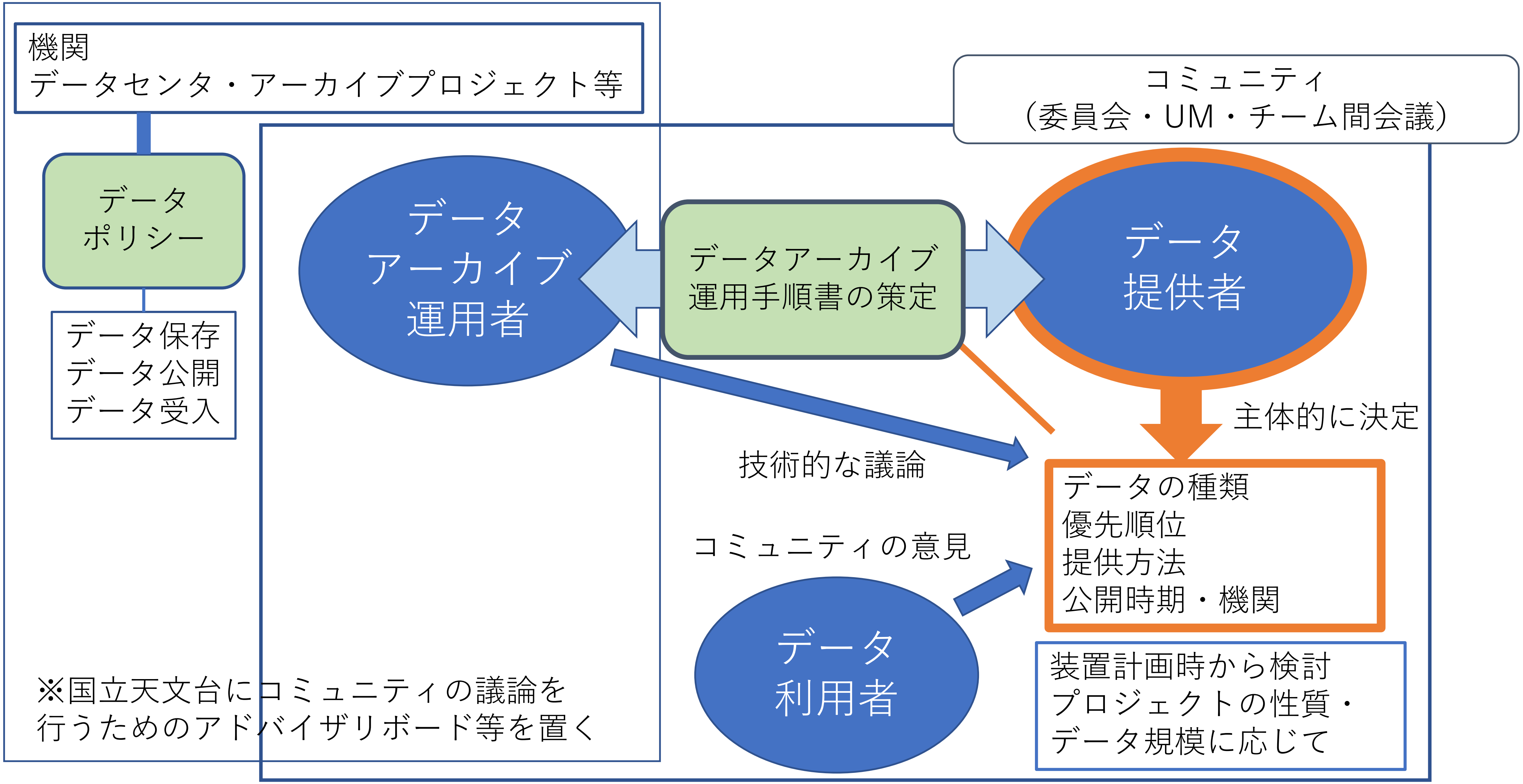
- データ利用者（業務にかかわらずデータを使う人）
  - **サイエンス・エンジニアリングを効率的・発展的に行える**
    - 様々なデータを品質管理された状態で便利に検索して利用
- データ提供者（装置，観測PI，データ処理者等）
  - データ保全場所と成果検証の手段が提供できる
  - **想定された以上の**データ価値・**装置/観測成果の向上**が望める
  - データの環境情報や品質情報などを散逸させず管理できる
  - データ管理にかかる負担を軽減して，**装置・サイエンスにより集中**できる
- データアーカイブ運用者（観測所・データセンタ・諸大学のアーカイブ）
  - **データ価値向上・サイエンス支援の向上により集中できる**
  - **長期に**データを**安定保存・公開**，**適時**の公開と**サイエンス促進**できる
  - **コミュニティの成果はデータアーカイブの評価 →正のフィードバックに**

データ利用・運用現場の各立場の皆さんへの提案

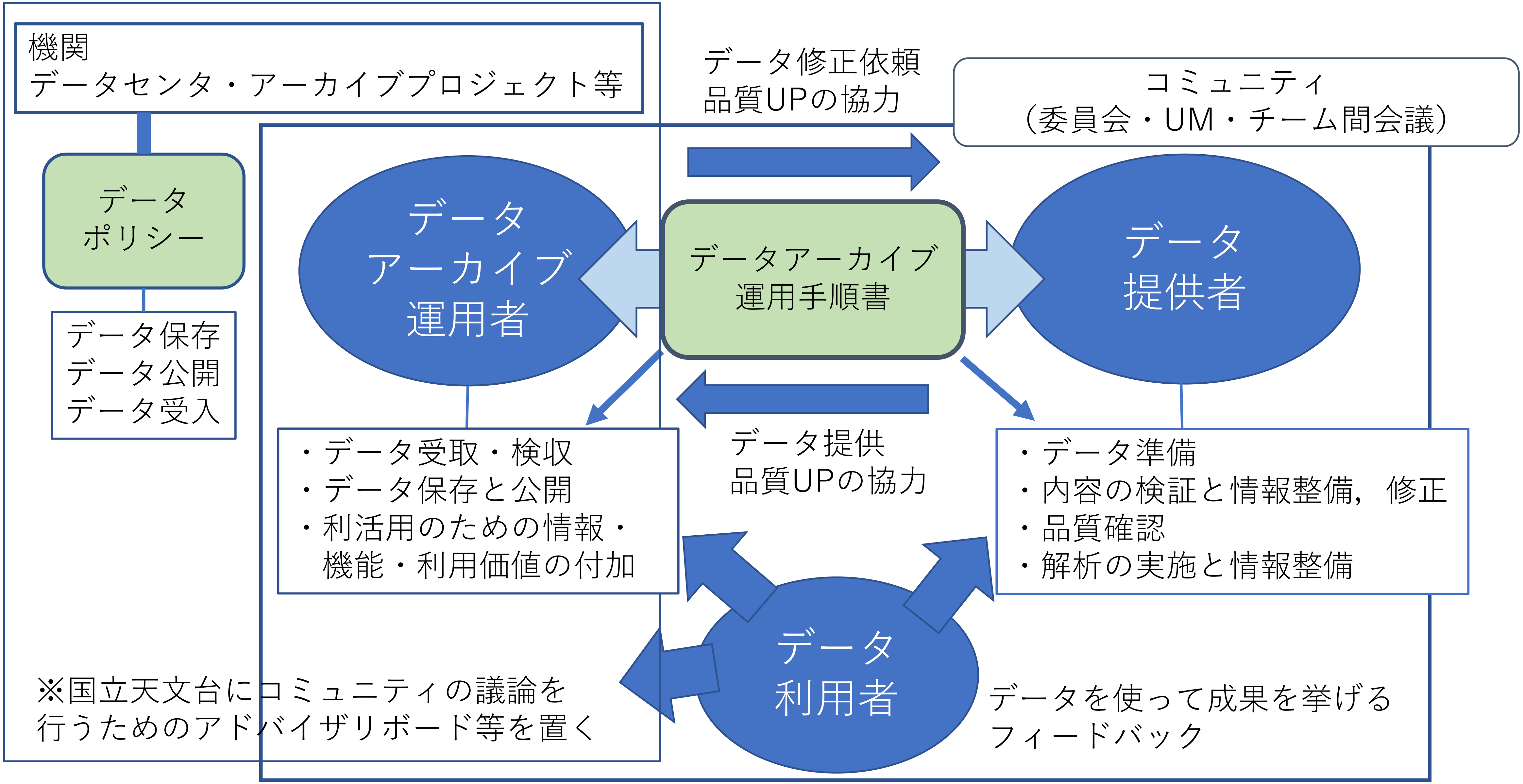
- **データ利用者**（データを使って研究，教育，開発などする人）
  - データを活用してどんどん成果を挙げる
  - データ公開・改善に必要な情報，意見を提供する（観測ログ取り，UMでの議論など）
- **データ提供者**（観測所，装置チーム，観測PI，データ処理者）
  - 運用手順書を作り可能な範囲でデータ公開に努める
  - データを整備・修正しアーカイブに提供
  - データ品質を向上する
- **データアーカイブ運用者**（観測所，諸研究機関のデータアーカイブ担当）
  - 運用手順書を作りデータを継続的に保全し公開する
  - データを検収する
  - データ品質と利便性，データ利用支援サービスの向上に努める

このように協力することで，データ利用者，データ提供者，データアーカイブ運用者の三者いずれにとっても最小限の負担で，データアーカイブからの利益を最大に得ることができ，今後の体制拡充にもつながる

DA運用改善モデル：データ公開設定



DA運用改善モデル：定常運用期



コミュニティ・機関等への提案

- **データ公開の重要性を共有**する
- データアーカイブの**目標と将来計画を共有**する
- 長期的貢献・社会還元の観点を持つ
- **生データ・科学プロダクト双方の品質が必要**
- 運用の負荷と貢献を**正当に評価・サポートする**
  - **→人材育成**（研究・技術面・国内外を俯瞰し，データ運用をリード）
- 大学等研究機関と大学共同利用機関が**双方向に協力**する
- **国立天文台に期待されること**
  - **長期的にデータを集約し共同研究の拠点**となる
  - 日本のデータアーカイブ運用の**方向性を議論する委員会を継続的に置く**
- **社会全体のサポートを得る努力**をする
  - オープンサイエンス化や成果の社会還元など

実際の運用はどうなるのか

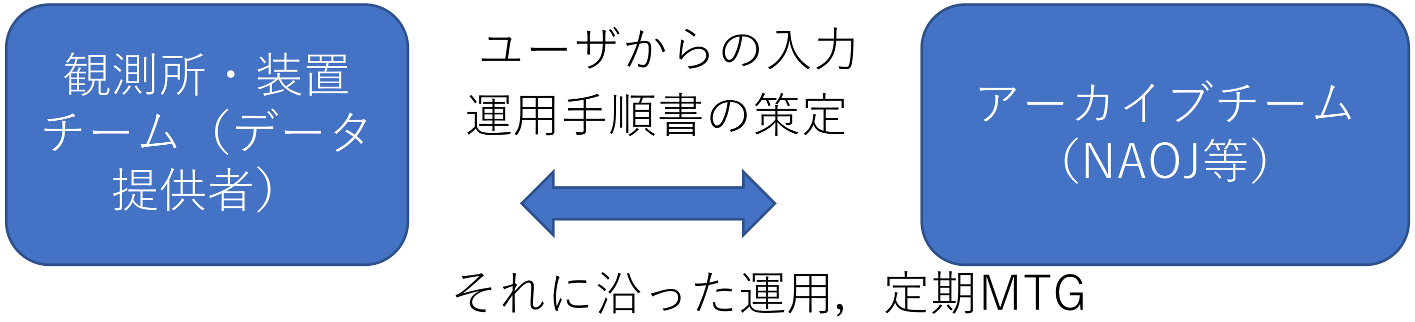
- 技術的，科学的な要請は別途きちんと議論して計画書や運用手順書に落とす

Q.　既存観測所・装置チームが何を改善すればよいのか

A.　・観測所チーム・アーカイブで相談して，アーカイブするために確認をするデータフォーマットや値の範囲，修正方法などについて運用手順書を作り，それに沿う．また，定期的に情報交換する．

・アーカイブ機関はこのようなやり取りを継続運用するためのデータポリシー，データ受け入れポリシー，委員会を作る

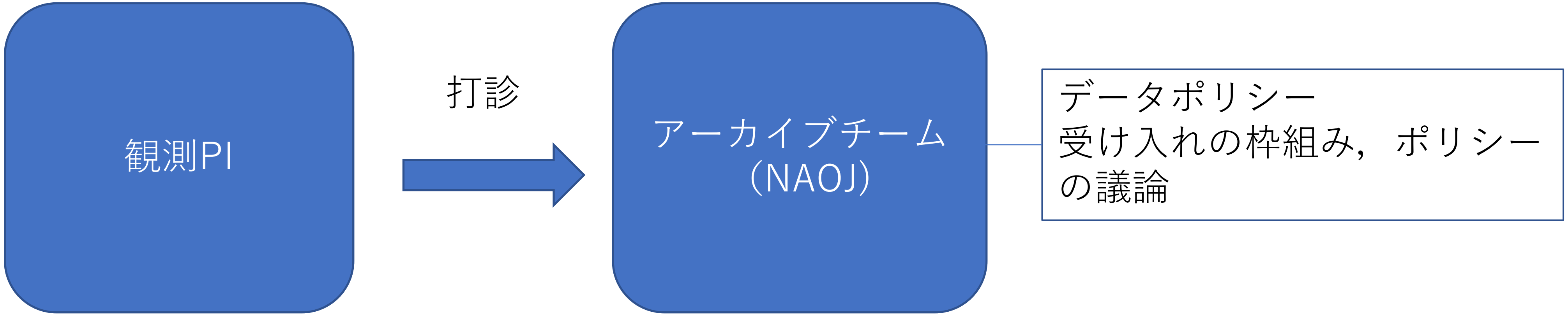
・ユーザの意見を取り入れる枠組みを作る



実際の運用はどうなるのか

Q. 個別の観測PI・大学研究室がデータを公開したいときどうしたら良いのか

A. ・自機関でホストしても良いが，難しい場合，国立天文台の既存アーカイブチーム（天文データセンター）に相談し，実現できる方法を探る
・長期的・戦略的には国立天文台にデータを集約・リンクすることを目指す
・そのようなデータ受け入れの窓口を国立天文台に持てるよう台内やコミュニティで検討することが望ましい（データポリシーの策定，継続的な委員会を設けること）



この提言をもとにした，次の議論とは.

- **技術面**
  - データアーカイブの検索・取得・解析に**期待される機能とその実現方法**
    - 統合的なデータ検索（バックエンド，メタデータ整備，VOプロトコルの活用など）
    - サイエンスプラットフォームなど
  - **信頼性のあるデータを作る**ため，将来計画を進めるために必要な**データ解析・較正の実現方法とソフトウェア開発・それを進めるための体制**
- **体制・ルール作り**
  - データポリシーの策定と**データ公開の受け皿**の整備，運用手順書策定
  - **コミュニティ議論**の継続的な受け皿づくり（委員会など）
  - サイエンス・開発運用両面での**データ科学拠点の醸成と人材育成**
    - 可視赤外を越えた多波長・他分野横断のデータアーカイブ運用検討 (IPAC的?)
    - 有効な国際協力・共同研究のデータ運用方法
      - ミラー・再配布・再解析・Interoperability実現など
    - コミュニティしてコヒーレントな解析・アーカイブ開発体制，解析体制
  - **予算・人材確保**の方法・

提言へのご意見をお願いします．

- 趣旨・全体
- 意義（２－３章）
- 運用提案（４，５章）
  - アーカイブされないデータを利用可能にする作業（写真乾板など）
- 体制提案（6章）
- 提言書へのフィードバック

ここからの進め方

- 声明採択の予備投票を行う
- 可能なフィードバックを9月14日(火)中までに提言に反映する
- 9月15日(水）の総会にて声明採択の決を取る

Zoomによる予備投票

- 提言の内容にご賛同いただけますか
- ご賛同が十分に多ければ来週の決議に進ませていただきたいと思います.

- 賛成
- おおむね賛同
  - この議論を踏まえ改訂されれば賛成の方，も含む
- 反対
- 光赤天連会員ではない

補足スライド

より良い＆持続的なデータアーカイブ運用にむけて データ提供者・アーカイブ運用者の立場から

- データアーカイブ運用は**業務の側面**があるが，**それ自体が面白い研究・開発，自身のサイエンスにリンクするもの**でもある
- 提言の実現には，**自発的な興味・動機とその維持**が必要である
  - でなければ若手も育たないし，研究者キャリアとして顧みられなければそもそも人は入って来ない
  - そういった魅力を持つデータアーカイブ運用であるべき
- **盤石なデータ運用＋研究者として，より発展的なデータ利活用・サイエンスを追求したい**
- データアーカイブ運用が**装置開発や科学計画の重要な一部として議論される**ことが必要
- **アーカイブサイエンティストが将来計画に参画してその責任を果たし，コミュニティは継続的なデータ公開に向けた責任を果たす**
- **データ解析手法を開発する・解析を実施し共同研究・コミュニティのためにデータを作る作業**も広い意味で**データアーカイブ運用**（データ価値を決める鍵のひとつ）である

三者の役割と協力方法　（5章）

（1）データ利用者の役割

1） 観測データを積極的に利用して科学活動を進め成果を挙げる

2） 委員会やデータアーカイブのユーザーズミーティング等を通して公開データの決定や公開方法への意見を入力する

3） データアーカイブ運用に必要となる情報（観測を実施する時には観測ログや較正データ取得への協力なども含む），また，アーカイブデータの利用を通して得られたデータ品質やアーカイブ機能への希望などについての情報をデータ提供者およびアーカイブ運用者へ提供（フィードバック）し，データアーカイブの利用価値向上に協力する

(2) データ提供者の役割

1) 提供データの種類，優先順位，提供方法，公開すべき時期と期間を決定する．新規開発の装置など大規模なデータアーカイブの場合には，コミュニティとの議論を行った上で判断を下すことが望まれる．

2) 提供するデータを作成する(具体的な手順はアーカイブ運用手順書に記述する)
   - データアーカイブ運用者に提供するデータの準備
   - データの内容の検証と整備
     - データが「アーカイブ運用手順書」に定義されたフォーマットを満たし利用可能であることの確認を行う
     - アーカイブするデータのヘッダ，時間・座標・観測条件を含むメタデータの検証と整備を行う
   - 品質確認
     - データが「アーカイブ運用手順書」に定義された品質を保持していることの確認を行う
   - データ解析の実施，データ解析ツール・解析／較正方法の提供
     - データ解析や較正に必要な情報の整備を行う

3) データアーカイブ運用者へデータ（生データまたは処理済みデータ）を提供する

4) アーカイブ運用者と協力して継続的にデータ品質の向上に努力する

5) アーカイブ後のデータにアーカイブに適さない不備が見つかった場合は修正を行う

6) 以上は大型プロジェクトや観測所・装置プロジェクトでは実施を推奨するが，データ規模，プロジェクト規模に応じて柔軟に実装可能な手続きを策定する

(3) データアーカイブ運用者の役割

1) データ提供者との協議により，アーカイブ運用手順書を策定する

2) データの受け取りと検収

- データ提供者からデータを受け取り，データフォーマットや内容が「アーカイブ運用手順書」に従っており，アーカイブ可能な状態であることを確認し，必要に応じてデータ提供者へ修正を依頼する

3) データの保管と公開

- 上で述べたデータアーカイブの要件とアーカイブ運用手順書に従い，データを保管し公開する

4) データ利活用のための情報・機能・利用価値の付加

- 上で述べたデータアーカイブの要件やデータ提供者との協議を指針として，継続的に，公開データの品質・利便性の向上やデータ利活用のためのサービスの充実（データ検索・取得機能，表示機能，解析機能等の科学研究のサポート機能を含む）に努め，公開データに科学的・教育的・社会的価値を付加する作業を行う
- 公開データの内容やアーカイブ機能に対するユーザーの希望を適宜受け付け運用方法に反映する機会（ヘルプデスク，ユーザーズミーティング等）を持つ

アーカイブ全体像と体制への提案（6章）

1) 光赤外線コミュニティのメンバーは，観測データの保全と利活用を進めることで科学推進に資することを目的に，大学・研究機関等と共同利用機関が双方向に協力してデータアーカイブ運用を行うことの重要性を理解する．

2) 大学・研究機関等の個々の観測プロジェクトは，積極的なデータ公開が装置開発や観測実施による科学的成果を増大し，観測データの価値や信頼性を高めることの重要性を認識する．また，短期的な特定の成果獲得のためのデータ利用に閉じることなく，長期的な天文学研究への貢献，及び社会的還元の観点から，積極的にデータアーカイブ運用へ貢献・協力する．

3) データアーカイブの有用性は，提供されるデータの品質および信頼性と，データ検索やダウンロード等のデータ利用支援機能の両者が揃って初めて最大化される．データ提供者においては，生データの整備に加えて，より科学的利活用に直結する処理済みデータの公開も望まれる．ただしその品質は，適切な解析・較正処理を行うことの重要性はもちろんのこと，生データアーカイブの完全性と品質に根差して初めて検証可能で信頼できるものとなることに留意する．データアーカイブ運用者においては，データ提供方法や価値付加の工夫を行いデータの再利用を促進することで，データ利用者から公開データの品質向上に役立つ情報のフィードバックを積極的に取り入れデータ提供者と共有すべきである．課題によっては，データ提供者と協力してデータや較正情報の取得方法など装置・観測に立ち返った改善も含めて取り組むことが必要である．また，そのようなデータアーカイブ運用を前提とし，データ利用者も含めてデータの品質確保に必要な観測ログなどの情報取得を協力して継続的に行うことが望まれる．

4) 天文コミュニティで共有する需要・目標に即してデータアーカイブ運用が進められることが理想的である．データアーカイブを構築・運用しようとするチーム・組織は，天文学コミュニティとの協議を通し，必要な人的・計算機・設備リソースなどを含めて天文学コミュニティの将来計画に根差したデータアーカイブ運用計画を策定する．

5) 大学共同利用機関であり日本の地上光赤外線天文学コミュニティおよびその共同研究推進の中核となる国立天文台は，宇宙科学研究所等他の機関とも協調することで，大学等観測研究プロジェクトのデータアーカイブ活動を支援し，継続性をもって日本全体の天文観測データを集約し，国内外の将来観測計画と連携することで，それらデータを用いた日本の共同研究推進の拠点となることが期待される．またそのための十分な人員体制と計算機リソースを確保し，人材の教育・育成を継続して行っていくことが望まれる．

6) 国内の天文データアーカイブ活動を俯瞰し，その方向性を定め，適切な助言を行うアドバイザリボードを国立天文台に継続的に設置することを強く望む．このボードには，データ提供者，アーカイブ運用者，およびデータ利用者それぞれの立場を理解する有識者が参加すべきである．このボードでは，現行のデータアーカイブ運用とデータ利用者の橋渡しをするとともに，大規模化していくデータに対する科学的な優先順位付けなど含め，日本全体のデータアーカイブ運用の指針が議論されることが期待される．

7） 日本全体のデータアーカイブ運用体制の維持と発展のためには，データアーカイブ運用に関わる組織の執行部やコミュニティ全体から，天文学推進におけるデータアーカイブの重要性が認識されることが重要である．同時に，データ提供者およびデータアーカイブ運用者両者の，データ運用にかかる負荷やその天文学推進への貢献が正当に評価され，予算的なサポートや関係者がデータアーカイブ運用を継続し人材を育成するためのインセンティブにつながるようにすることが重要である．

8） 基礎科学研究にあてられる予算が逼迫する状況下では，上で述べたようなデータアーカイブ運用体制の改善を進めると同時に，観測データアーカイブの価値に対して天文学コミュニティにとどまらず社会全体のサポートを得ることが重要である．そのためにも，天文学コミュニティ全体でデータアーカイブ運用の長期の方向性を共有し，オープンデータ・オープンサイエンス化への対応を進めるとともに，アーカイブデータを活用した科学的成果の社会還元に向けて積極的に情報を発信していく．