

大規模天文データ科学時代に むけたデータアーカイブ の構築

2020.2.14 宇宙科学情報解析シンポジウム@宇宙研

古澤 久徳@国立天文台（天文データセンター・ハワイ観測所）

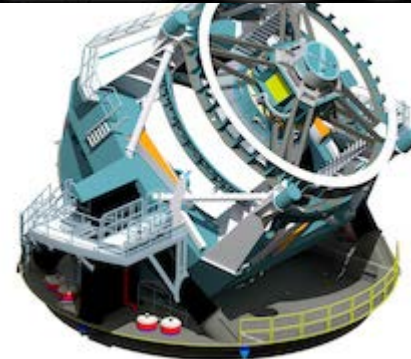
for データアーカイブワークショップ有志

技術やサイエンスの面白い話でなくて
すみません。。



天文ビッグデータ時代とデータ運用

- ▶ HSC/PFS/Euclid/WFIRST/LSSTなど大サーベイ観測
- ▶ 処理済みデータを戦略的に活用する巨大データベースプロジェクト
- ▶ データをどう作り、どう保存し活用するのが、成功の鍵であるとともに責務



データアーカイブの重要性と困難 すばる・HSCの経験から

- ▶ 有効なデータ活用
 - ▶ 確実なデータの保全と（再）利用のための整備・価値付加
→ 運用に根差した活動の積み上げ
- ▶ 日本の可視近赤外分野のデータアーカイブ体制
 - ▶ まだ十分とは言えない・場所により危機的状況
- ▶ 観測データ運用を必要とする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に
 - ▶ 日本のデータアーカイブの目的の再確認
 - ▶ 目指すべき方向性と実現のための努力の方法の共有

政策と共同利用機関国立天文台の状況

- ▶ 政策や国研の体制的な変化
 - ▶ オープンデータ・オープンサイエンス推進
 - ▶ 大学の自助活動と大学共同利用機関の在り方（4機構連合など）
 - ▶ 第4期中期目標：大学共同利用機関→貴重な学術データ等の保有提供、全国的な視点
 - ▶ 対して、運営費交付金の継続的な減少
 - ▶ 一方、国立天文台（共同利用機関）の状況
 - ▶ 小体制ながら継続性をもって光赤外データの集約拠点として機能してきた
 - ▶ 予算バランスの変化（運交金・フロンティア経費）→ 選択と集中の進行・効率化！
 - ▶ データ専門委員会：国立天文台観測データの扱いについて提言してきた
 - ▶ データポリシー(2014)、 データ利活用の提言・答申(2016,2018) →十分な実現が困難。。
- 日本全体のデータアーカイブをどう作り管理すべきなのか皆で考えたい

データアーカイブ ワークショップ

- ▶ 光赤外線分野のデータ運用者有志によるキックオフの議論
- ▶ 2020.1.29-30 (1.5日) 約25人参加
- ▶ データアーカイブ
 - ▶ 取巻く環境・背景
 - ▶ 存在意義
 - ▶ 維持・向上の方向性

データアーカイブワークショップ

[トップページ](#)

[開催主旨](#)

[プログラム](#)

[参加者](#)

ワークショップ開催主旨

目的

近年、大学・公開天文台の望遠鏡・装置開発が活発に行われ、同時に大規模観測プロジェクトの推進が進む中、日本のデータアーカイブの置かれる状況を整理して理解し、データアーカイブの意義・在り方を確認した上で、その維持・向上のため向かうべき方向について意見交換を行う。本ワークショップは、その議論のキックオフのための検討会の位置づけであり、各大学所有の観測装置やすばる望遠鏡から得られるデータの運用といった具体的な課題を持つ光赤外分野の有志を中心に課題をあぶり出し、次回以降のより全体的なデータアーカイブの在り方に向けた検討会で議論を深めたい。

議論のテーマ

データアーカイブを取り巻く背景整理

日本・国政策・大学共同利用体制・国立天文台の状況
世界の時流

データアーカイブの存在意義

何をするためのものか
それが日本・世界の天文学にとってどのような位置づけか

データアーカイブの維持・向上の在り方

何を・どこで保持し、どのように提供し、何をしたいのか、誰が行うのか

世界のデータアーカイブの状況

- ▶ 調査：主要データアーカイブの運用・開発のコアの人
重要なこと、コミュニティ内のDAの位置づけの認識は？
- ▶ データアーカイブ会議@シドニー(2018.8) +
ADC有志で個人的に光赤外系DA担当者へ質問
- ▶ CADC、IPAC、ESA、ESOの担当者から返答
- ▶ 8個の質問
 - ▶ 1)データ種類・決め方、2)体制規模、
3)データ提供者との役割分担、4)データ優先付け、
5)計算機管理、6)最重要の役割、7)予算措置、
8)理解・サポート

世界のデータアーカイブの状況



- ▶ 役割
 - ▶ コミュニティのデータの集約と維持・科学指向・研究の支援
- ▶ 保持するデータ・優先度
 - ▶ コミュニティの全データ・優先度なし・処理済みデータに重心
 - ▶ 親組織・コミュニティの意見を踏まえて決める
- ▶ 体制
 - ▶ 1アーカイブに対して20-30人規模（DC全体で>100人規模）：NAOJ≤4人/アーカイブ
- ▶ データ提供者との役割分担
 - ▶ 提供者側：テスト、品質評価
 - ▶ アーカイブ側：テスト+修正依頼、利用価値付加に注力
- ▶ 予算・サポート
 - ▶ 問題ない：コミュニティの体制としてサポートされている高い論文生産率に裏付けられている

ワークショップ議論の概要

- 国立天文台SMOKA/JVO/HSC ..
- 国立天文台すばる望遠鏡
- 広島大かなた望遠鏡
- 京都大せいめい望遠鏡
- 東京大TAO
- 東京大木曾Tomo-e-Gozen
- 宇宙研あかり
- 宇宙研DARTS・C-SODA
- 東京工業大MITSuME
- 西はりま天文台なゆた望遠鏡
- ほか

大学・公開天文台の状況・課題

▶ 意見をいただいた機関

- ▶ すばる、東大(TAO, Tomo-e-Gozen)、京大(せいめい)、広大(かなた)、宇宙研DARTS・AKARI、西はりま、東工大MiTSuME など
- ▶ データは保存公開したい。大学自身で整備する動きもあるが（広大、）
- ▶ （特に長期の）保存公開はNAOJへの期待（Tomo-e、せいめい）
 - ▶ 必ずしもすぐ論文出版されるデータばかりではない
- ▶ データ整備はしっかりしたいが、人不足・短期雇用（SMOKA、HSC、。。）
- ▶ 装置開発中はデータ整備の検討が思うように進まない
 - ▶ 設備お金もない（せいめい）
- ▶ 処理済みデータ公開による科学促進を（すばる、Tomo-e-Gozen、TAO）
- ▶ 特色あるデータで社会・教育など戦略的な予算の誘致を（Tomo-e）
- ▶ 宇宙ミッションでは一定のデータ運用ポリシーが機能。
系統的なデータ解析体制が課題（DARTS・AKARI）

データアーカイブの意義・役割

- ▶ 必要性
 - ▶ **必要**
- ▶ 保管と再利用の2大目的
 - ▶ 再現できないデータの**保全**
 - ▶ 将来の解析・未知事象の発見を可能とする
 - ▶ 科学結果に対する**エビデンス**保全
 - ▶ **科学促進** → 処理済みデータ
- ▶ **公開することが重要**
- ▶ データは人類の文化遺産という考え方が求められる

どのようなデータをアーカイブするのか 誰がそれを決めるのか

- ▶ 基本はデータ提供者が価値あるかどうか判断する
 - ▶ 将来の巨大生データの保全については議論がある
- ▶ 関係するコミュニティの意見を聞いて決定する
- ▶ 優先順位付けがあるならばそれも同様
- ▶ それを受容できるアーカイブ→
ミッションの定義（e.g., 宇宙研） ・ データ運用ポリシー
- ▶ 処理済みデータ ・ 解析情報の提供が望ましい

どのようなデータをどのような状態で

- ▶ **使える状態**で公開
 - ▶ 解析できること（広い意味で？）
 - ▶ 取得時の情報があること
- ▶ 品質評価
 - ▶ ベストエフォートで良い。**公開を遅らせない。**
 - ▶ 間違いは修正されていく
- ▶ 必要なデータを探しやすいこと

データ提供者とアーカイブの役割分担

▶ データ提供者

- ▶ **検証・整備**（ヘッダ、時間・座標・観測条件などメタ情報）

- ▶ **品質確認**（公開を遅らせない）

- ▶ **解析・解析ツール**（共同利用）・少なくとも**解析情報**の整備

▶ アーカイブ

- ▶ 検証確認・修正の依頼

- ▶ **保管と公開**作業

- ▶ **データ活用のための情報・利用価値の付加**

▶ 進め方・体制

- ▶ ルール・インターフェースを決めて**守る**

- ▶ **設計段階から**どう作り保存するのか**装置・アーカイブ**で相談

日本のデータアーカイブの体制は

- ▶ 日本全体の観測データの集約・保全を目指したい
 - ▶ タイムドメインなど過去のデータを一望したい
- ▶ 共同利用機関（国立天文台+）には
有期のプロジェクトではなく**継続性**、
全国的な視点に立って**データの集約拠点**の役割を期待
- ▶ 30年規模の長期ビジョンを持つ

組織・コミュニティのサポート獲得

- ▶ 研究会でのアーカイブデータによる成果発表の促進
- ▶ 講習会でのアーカイブデータ利用による意識付け
- ▶ 装置開発者の動機付け
 - ← データ公開が望遠鏡・装置の価値を高めることへの理解
- ▶ 科学雑誌・外部資金によるデータの公開要請
オープンデータへの対応を進める
- ▶ 社会・教育普及利用・天文コミュニティにとどまらない視点
で議論しサポートを得る
- ▶ 長期ビジョン・ロードマップを持ってビッグピクチャーを語る
- ▶ 成功例・成果がみこめるケースをまとめる
 - 文書としてきちんと残す

天文データは人類の文化遺産、 だからどうするか

- ▶ 「今」の天文学の発展のためのアーカイブ
→ これは本当に大事
- ▶ データを「後世」に残すべき
→ 今の天文学も信頼あるものにする
- ▶ 少なからず研究人生を投入する & 持続のためには
正しい動機付けとサポートが必要

しかしここで、データベースを活用してその時代時代の天文学の研究を進めることと、そのデータベースを文化遺産として後世に伝えることとはほとんど別の話であることに注意しなければならない。100年前の天体写真乾板は保存庫からとり出せばすぐに見ることができる。し

後世に残すべき質のデータを大量に生み出すに違いない。すばるのデータをはじめさまざまなデータベースを、時代時代の技術進歩に合わせて更新し、文化遺産として後世へ伝えてゆくには、技術開発もさることながら多くの資金、労力も必要であろう。地味ではあるが、何十年後に大きな意味を持つこのような仕事は、国立天文台にしかできない重要な仕事の一つではなかろうか。その成功は多少大げさに言えば、文化の継承における新しい形態を生み出すことにもなるのである。

S. Okamura

ISSN 0915-9770

国立天文台

天文学データ解析 年報
計算センター

第4号

(1992年4月～1993年3月)

宇宙の姿は残せるか

岡村定矩 (東京大学大学院理学系研究科天文学専攻)

まとめ

- ▶ 日本のデータアーカイブの今後の維持発展のために議論を始めた
- ▶ 3月下旬にアーカイブデータで仕事をしているユーザの意見を聞く会を計画中
- ▶ 1回目WSの資料に興味のある方は古澤まで
furusawa.hisanori@nao.ac.jp
https://www.adc.nao.ac.jp/people/~furusawa/work/da_ws/