



国立天文台の

# 天文アーカイブを取り巻く状況について

高田唯史(国立天文台天文データセンター)

# 今日の話の内容

- 国(文科省など)による施策との関係
  - ✓オープンデータ、オープンサイエンス
  - ✓4機構連合体構想(大学共同利用機関のあり方に関する議論)
- 国立天文台のアーカイブに関する近況
  - ✓予算バランスの変化(運営費交付金とフロンティア経費)
  - ✓(観測データ関連)計算機システムに関する議論(効率化、特にコスト面)
  - ✓データアーカイブについての議論
    - ◆台外機関のデータの関する考え方(MoUベースの運用、データ受け入れ)
    - ◆データの階層化(重要度だったり緊急度だったり)の考え方
    - ◆オンプレ?クラウド?(明日の議論でもう少し詳細を紹介できれば)
    - ◆生データと処理済みデータ+そこからの成果に対する考え方

# 国の施策に関する事

## • オープンデータ・オープンサイエンス

- オープンサイエンスとは(国情研オープンサイエンス基盤研究センター(<https://rcos.nii.ac.jp/openscience/>)より抜粋)

デジタル時代に鑑み、これまで以上にオープンで、多様な可能性をもって行うことができるようになった研究活動の諸側面の総称

- ✓ サイエンスはよりオープンであるべきという理念
  - ✓ 説明責任や透明性などの観点(主に行政サイドから来るが、自然科学においても当然要求されるものとなっている。)
- オープンデータとは?
    - ✓ 特定のデータが、一切の著作権、特許などの制限なしで、全ての人が望むように利用・再掲載できるような形で入手できるべきという考え方

◆ 国の施策としてデジタル時代に備えた対応を組織的にしていこうとする動きは着実に進み、様々な方面での議論が行われているとともに、組織体系にも影響が出始めている。天文データアーカイブもその流れの中に確実に入っていくことになる。

- ✓ オープンサイエンスの推進について(文科省・科学技術・学術審議会総合政策特別委員会 H28.11.24)  
([https://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu22/siryo/\\_icsFiles/afieldfile/2016/12/08/1380241\\_04.pdf](https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu22/siryo/_icsFiles/afieldfile/2016/12/08/1380241_04.pdf))
- ✓ 内閣府及び日本学術会議におけるオープンサイエンスに係る検討状況  
([https://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/040/attach/1413786.htm](https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/040/attach/1413786.htm))
- ✓ オープンサイエンスの深化と推進に関する検討委員会(日本学術会議)  
(<http://www.scj.go.jp/ja/member/iinkai/openscience24/openscience.html>)

# 大学共同利用機関に関する考え方

- 第4期中期目標期間における大学共同利用機関の在り方について(審議のまとめ)  
([https://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/010/toushin/1412585.htm](https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/010/toushin/1412585.htm)) 2018□12□14□(□□□□□□□□)

「個々の大学では整備・運用が困難な最先端の大型装置や**貴重な学術データ等**の研究資源を保有し、これらを全国的な視点に立って共同利用・共同研究に供していること」

国立天文台における天文データアーカイブは、この観点について一定の役割を期待されているとは思われるが、その範囲がどこまでなのかの議論はあまりちゃんとされていない。

国立天文台が自ら生産するデータについては、国立天文台天文データ専門委員会により策定された答申は現台長には提出された。

データポリシー提案

([https://www.adc.nao.ac.jp/J/cc/public/center/ADC\\_Data\\_policy\\_20140522.pdf](https://www.adc.nao.ac.jp/J/cc/public/center/ADC_Data_policy_20140522.pdf))

公開データ受け入れポリシー

([https://www.adc.nao.ac.jp/J/cc/public/center/ADC\\_Data\\_acceptance\\_policy.pdf](https://www.adc.nao.ac.jp/J/cc/public/center/ADC_Data_acceptance_policy.pdf))

「大学共同利用機関として備えるべき要件」については、主に以下のような内容が考えられるところであり、今後、文部科学省において、科学技術・学術審議会の意見を聴き法令等において具体的に定めることが必要である。

- 開かれた運営体制の下、各研究分野の研究者コミュニティ全体の意見を取り入れて運営されていること
- 各研究分野に関わる大学や研究者コミュニティ全体を先導し、最先端の研究を行う中核的な学術研究拠点であること
- 国際的な学術研究拠点として、各研究分野における我が国の窓口としての機能を果たしていること
- **個々の大学では整備・運用が困難な最先端の大型装置や貴重な学術データ等の研究資源を保有し、これらを全国的な視点に立って共同利用・共同研究に供していること**
- 時代の要請や学術研究の動向に対応して、新たな学問分野の創出や発展に戦略的に取り組んでいること
- 優れた研究環境を生かした若手研究者の育成に貢献していること

# 4機構＋総研大連合体構想

(第4期中期目標期間における大学共同利用機関の在り方について(審議のまとめ)より)

4機構:自然科学研究機構、高エネルギー加速器研究機構、情報・システム研究機構、人間文化研究機構

## • 運営の効率化に向けた取組

- 各大学共同利用機関法人が、これまで各々で蓄積してきた技術・経験・ノウハウを持ち寄り、共同で取り組むことで効率化が見込まれる業務(例えば、広報、IR、評価、施設・設備のマネジメント、調達・契約、法務、知的財産男女共同参画に係る取組、研究不正への対応、情報セキュリティ職員の研修、産業界との連携・地方貢献活動に係る窓口の設置等)を実施する。

## • 研究力の強化に向けた取組

- 研究連携促進のための基本方針を策定の上、異分野融合による研究領域の拡大と新分野の創成に向けた研究プロジェクトを実施する。また、大学共同利用機関の国際化を促進するため、海外リエゾンオフィスや外国人研究者の相談窓口を共同して設置する等の取組を実施する。さらに、ポストドクターのキャリアパス支援等、若手研究者の育成に取り組む。

## • 大学院教育の充実に向けた取組

- 総合研究大学院大学における大学院教育に関して、基盤機関である大学共同利用機関が有する海外の研究機関とのネットワークを生かして、国際共同学位プログラムを策定するとともに留学生のリクルート等を実施する

その中で、「連合体」で実施を検討している事業例－研究力強化の1つとして、「分野を横断する共通知であるデータサイエンスの推進(仮)」というものが持ち上がってきている。

([https://www.mext.go.jp/kaigisiryu/2019/11/\\_icsFiles/afieldfile/2019/11/22/1422630\\_006.pdf](https://www.mext.go.jp/kaigisiryu/2019/11/_icsFiles/afieldfile/2019/11/22/1422630_006.pdf))

但し、天文データセンターの名の下に書かれている記述は、センター長以下専任職員が感知していないという問題も、

# 「連合体」で実施を検討している事業例 – 研究力強化（3）

## 事業名

分野を横断する共通知であるデータサイエンスの推進（仮）

## 事業概要

連合体が主体となり、以下の事業を推進

- 情報・システム研究機構（ROIS）のデータサイエンス共同利用基盤施設（DS施設）等をコアとし、4機構傘下の各分野に於いて生産される多種・多大なデータの共有と解析を推進する「データセンター（仮称）」を組織整備し、分野横断的共通知としてデータサイエンスを推進する。
  - 各機関に於いて「データ集約センター」のような組織を作り、ROISのDS施設と連携させる。
- データサイエンスを総研大の共通科目とし、情報に明るい研究者の育成を図る。

## 検討状況

- これまで情報・システム研究機構では、機構傘下の各機関の持つデータを中心にデータサイエンスを展開しており、同時に人間文化研究機構と共同で文理融合課題に取り組んできたが、さらに本年度より自然科学研究機構・国立天文台の天文データセンターと連携し、観測データの処理・解析への情報処理手法の導入を開始した。
  - 今後、4機構傘下の機関や大学等とも連携し、幅広い分野のデータを対象とする予定
- 【工程】：上記のオープンデータ・オープンサイエンスに係る事業を21年度に見直した上で、「連合体」スタートと同時にその機能を移管し実施する。
- 【予算】：DS施設をコアとする「データセンター」の運営費は情報・システム研究機構が予算要求。これに連結する他機関の各機関にデータセンター機能を整備することについては、各機関の研究計画の中でどのように扱うかを検討し、予算化を図る。



# 国立天文台のアーカイブに関する近況(1)

- 予算バランスの変化(運営費交付金とフロンティア経費)
  - ✓ 運営費交付金は年々減少し、大型計画フロンティア経費(そのうちのそれなりの割合は補助金の形)が増大
  - ✓ 運営費交付金に対する考え方
- 計算機システムに関する議論(効率化、特にコスト面)
  - ✓ すばるFL位まではレンタルシステムで機能ごと調達し、天文台は少人数スタッフでまかなう形
  - ✓ だんだんと内製化が進んできた(レンタルと買取計算機のバランスの変化、非常勤職員増加)
  - ✓ データレートの高高速化などに伴いストレージ容量の急増(コストに直結)と画像処理演算にもHPC系の知識、人材が必要不可欠になったが、キャリアパスや給与の問題で解決困難
- データアーカイブについての議論
  - ✓ 台外機関のデータに関する考え方(MoUベースの運用、データ受け入れ)
    - 現執行部の主導で今後はMoUを結んだ相手からしかデータの受け入れをしないとの方針が出された
    - アーカイブのコストを誰が持つのか、運用に対する継続的な責任の所在の明文化などを求められる
    - 大学と共同利用機関との関係性についての認識のズレ?



# 国立天文台のアーカイブに関する近況(2)

## • データアーカイブについての議論(続き)

- ✓ データの階層化(重要度だったり緊急度だったり)の考え方
  - 今が旬のデータは高速アクセスできないと話にならない
  - 貯蔵コストを下げるにはより低速で容量大きな媒体で管理?
  - アーカイブは蓄積型で、それを使ったサイエンス・データ探索は(どのデータが頻繁にアクセスされるかの予想に基づく)データ階層化とマッチしない部分がある。
- ✓ オンプレ?クラウド?(明日の議論でもう少し詳細を紹介できれば)
  - クラウドにデータ置いた方が安い?という神話、一方で、大量の機器管理からの解放はあり得る
  - 使いたいときにすぐに使えるのか?安全性は?
  - 大きなデータを頻繁に扱うには向かない(コスト、ネットワークなどの影響)のでは?
- ✓ 生データと処理済みデータ+そこからの成果に対する考え方
  - 良く言われること:
    - 生データアーカイブはお金をかけるわりに論文出ない
    - もっと他の装置開発とか(新しい事)にお金をかけた方が成果が出る
    - 処理済みだけで良いのでは?(再現性や新たなキャリブレーションを導入する余地は?データの大きさ?)
- ✓ データ解析やデータアーカイブについての議論を天文コミュニティーと行う正式な場?
  - 天文データ専門委員会がなくなった影響

# 国立天文台データ解析・アーカイブ・公開システム

- 2018年3月1日より稼働したレンタル計算機システム
- 納入業者は富士通(株)
- 14個のサブシステムもしくはコンポーネント、そのうち、10個はアーカイブシステム
- 設置場所は主に三鷹だが、それ以外に、水沢、野辺山、岡山に設置
- 常駐SE3人、CEの常駐なし
- 翌営業日サポートが基本線(休日対応なし:法定停電時は別途契約)

多波長データ解析  
サブシステム

PCサーバ

PRIMERGY RX2530 M2 : 48台  
CPU総コア数 : 696  
総メモリ容量 : 8928GB



ディスク装置

ETERNUS DX100 S4 : 12台  
総ディスク容量 : 2484TB



ネットワーク  
ディスクアレイ

ETERNUS NR1000 F2650 : 1台  
総ディスク容量 : 126TB

大規模データアーカイブ・  
公開サブシステム

UNIXサーバ

SPARC M10-1 : 4台  
CPU総コア数 : 32  
総メモリ容量 : 256GB



ディスク装置

ETERNUS DX100 S4 : 33台  
総ディスク容量 : 9015TB

PCサーバ

PRIMERGY RX2530 M2 : 73台  
PRIMERGY RX2540 M2 : 4台  
CPU総コア数 : 1004  
総メモリ容量 : 7872GB



ネットワーク  
ディスクアレイ

ETERNUS NR1000 F2650 : 1台  
総ディスク容量 : 75TB



テープ装置

ETERNUS LT260 : 計1台  
総テープ容量 : 360TB



テープ装置

ETERNUS  
LT270 S2 : 計3台  
総テープ容量 : 3516TB

MASTARS  
SMOKA  
HSC  
ALMA  
野辺山  
VERA  
太陽  
岡山  
遠隔

【国立天文台情報ネットワークシステム】 ※ご提案対象外

ヴァーチャル天文台  
サブシステム

PCサーバ

PRIMERGY : 13台  
CPU総コア数 : 208  
総メモリ容量 : 1664GB



ディスク装置

ETERNUS DX100 S4 : 4台  
総ディスク容量 : 476TB

水沢地区データ解析  
サブシステム

PCサーバ

PRIMERGY RX4770 M3 : 1台  
PRIMERGY RX2530 M2 : 2台  
CPU総コア数 : 84  
総メモリ容量 : 320GB



ディスク装置

ETERNUS DX100 S4 : 1台  
JX40 S2 : 1台  
総ディスク容量 : 71TB



テープ装置

ETERNUS LT260 : 計1台  
総テープ容量 : 618TB

RISE  
測地解析  
水沢バックアップ

開発試験サブシステム

PCサーバ

PRIMERGY RX2530 M2 : 5台  
CPU総コア数 : 80  
総メモリ容量 : 1024GB



ディスク装置

ETERNUS DX100 S4 : 2台  
総ディスク容量 : 257TB

多波長  
JVO

共同利用機器群



PC

HP Z840 Workstation : 24台  
CPU総コア数 : 96  
総メモリ容量 : 1536GB

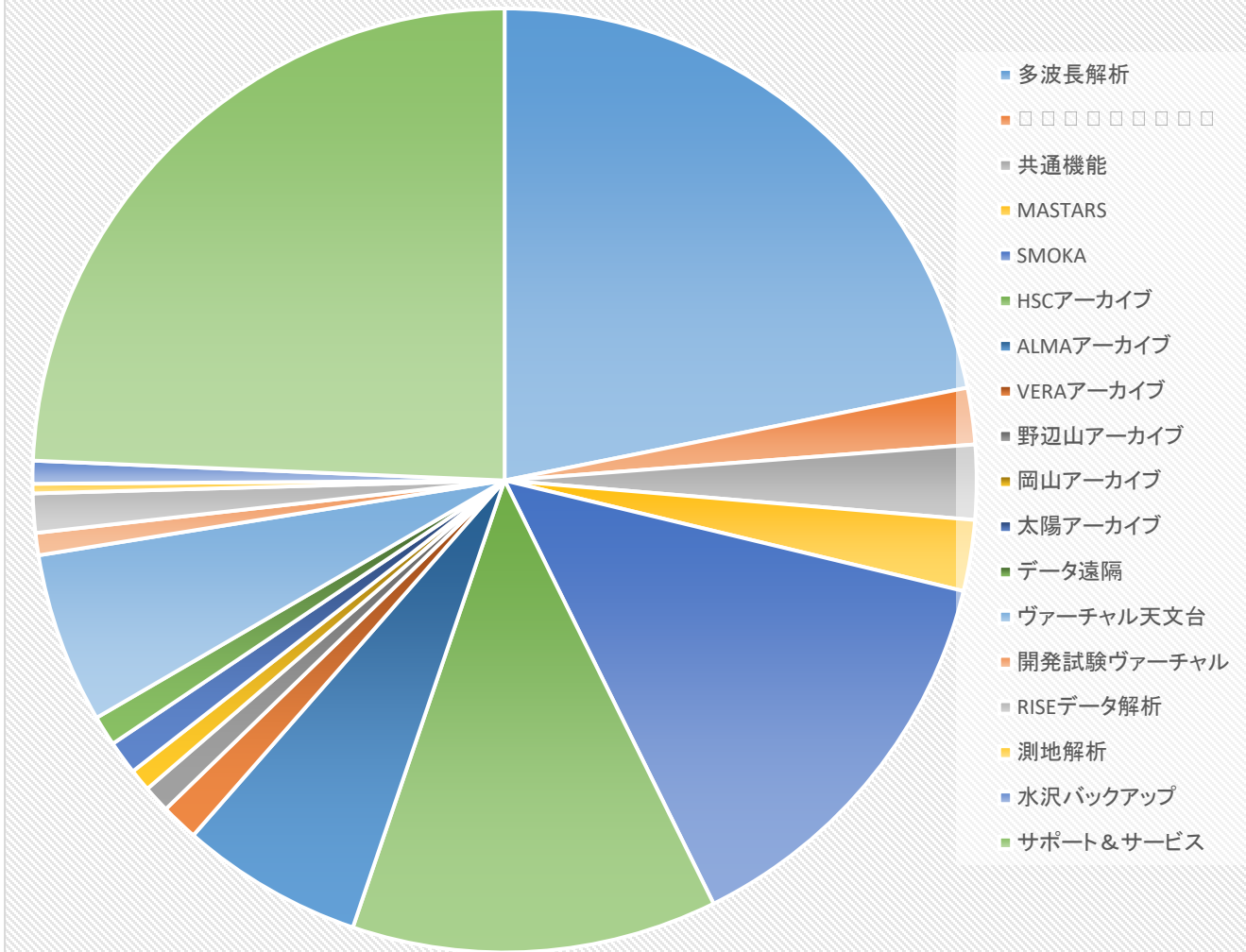
JVO

システム名	サーバー数	総コア数	総メモリ量 (GB)	総ディスク容量 (TB)	テープ装置 (TB) (非圧縮)
多波長解析	48	696	8928	2160	
□ □ □ □ □ □ □ □ □ □	3	48	768	189	
共通機能(端末など)	24	96	1536	72	
MASTARS	5	68	464	476	360
SMOKA	34	512	2720	2839	3000
HSCアーカイブ	23	340	3440	1925	516
ALMAアーカイブ	5	40	704	2261	
VERAアーカイブ	3	12	48	433	
野辺山アーカイブ	3	12	192	272	
岡山アーカイブ	2	16	128	204	
太陽アーカイブ	4	20	368	272	
データ遠隔	2	16	64	408	
ヴァーチャル天文台	13	208	1664	476	
開発試験ヴァーチャル					
RISEデータ解析	1	72	256	55	
測地解析	1	8	32	16	
水沢バックアップ	1	4	32		618

# レンタル計算機経費内訳

システム名	年額(円)	割合
多波長解析	43,157,724.4	0.215789
□ □ □ □ □ □ □ □ □	3,820,463.7	0.019102
共通機能	5,039,813.4	0.025199
MASTARS	4,798,781.8	0.023994
SMOKA	27,526,261.0	0.137631
HSCアーカイブ	24,710,020.6	0.12355
ALMAアーカイブ	12,344,873.4	0.061724
VERAアーカイブ	2,533,985.6	0.01267
野辺山アーカイブ	1,825,758.2	0.009129
岡山アーカイブ	1,511,966.5	0.00756
太陽アーカイブ	2,275,158.2	0.011376
データ遠隔	2,036,829.8	0.010184
ヴァーチャル天文台	11,554,424.9	0.057772
開発試験ヴァーチャル	1,502,505.5	0.007513
RISEデータ解析	2,641,661.5	0.013208
測地解析	640,648.5	0.003203
水沢バックアップ	1,548,684.4	0.007743
サポート&サービス	48,056,147.6	0.240281

2018年更新レンタルシステム費用内訳



# 買い取り機器予算(FY2013-FY2017)(単位 円)

	2013	2014	2015	2016	2017		合計
共同利用	6,565,182	3,150,459	7,562,118	1,618,530	5,619,415		24,515,704
SMOKA	2,850,541	2,587,356	3,580,327	4,091,314	4,728,149		17,837,687
HSC	2,380,938	632,556	0	4,968,000	485,222		8,466,716
JVO	2,762,558	0	0	0	0		2,762,558
ALMA	7,637,250	104,000	82,877	8,814,600	82,877		16,721,604
太陽	1,050,000	4,744,477	1,755,712	1,777,200	2,071,072		11,398,461
野辺山	0	6,048,000	264,171	1,139,760	264,171		7,716,102
水沢	0	648,188	909,360	0	0		1,557,548
MASTARS	0	0	172,800	0	0		172,800
合計	23,246,469	17,915,036	14,327,365	22,409,404	13,250,906		91,149,180

# ADC計算機関連使用電力と電気料金概算

- レンタルシステムについては、機器の構成も変わっていないので3月と7月の平均を取ったぐらいが平均的な使用電力と考えられる。
- 買取機器は3月から7月までに増設を行ったので、今後の2019年7月並の使用量であると仮定する。

(内訳)

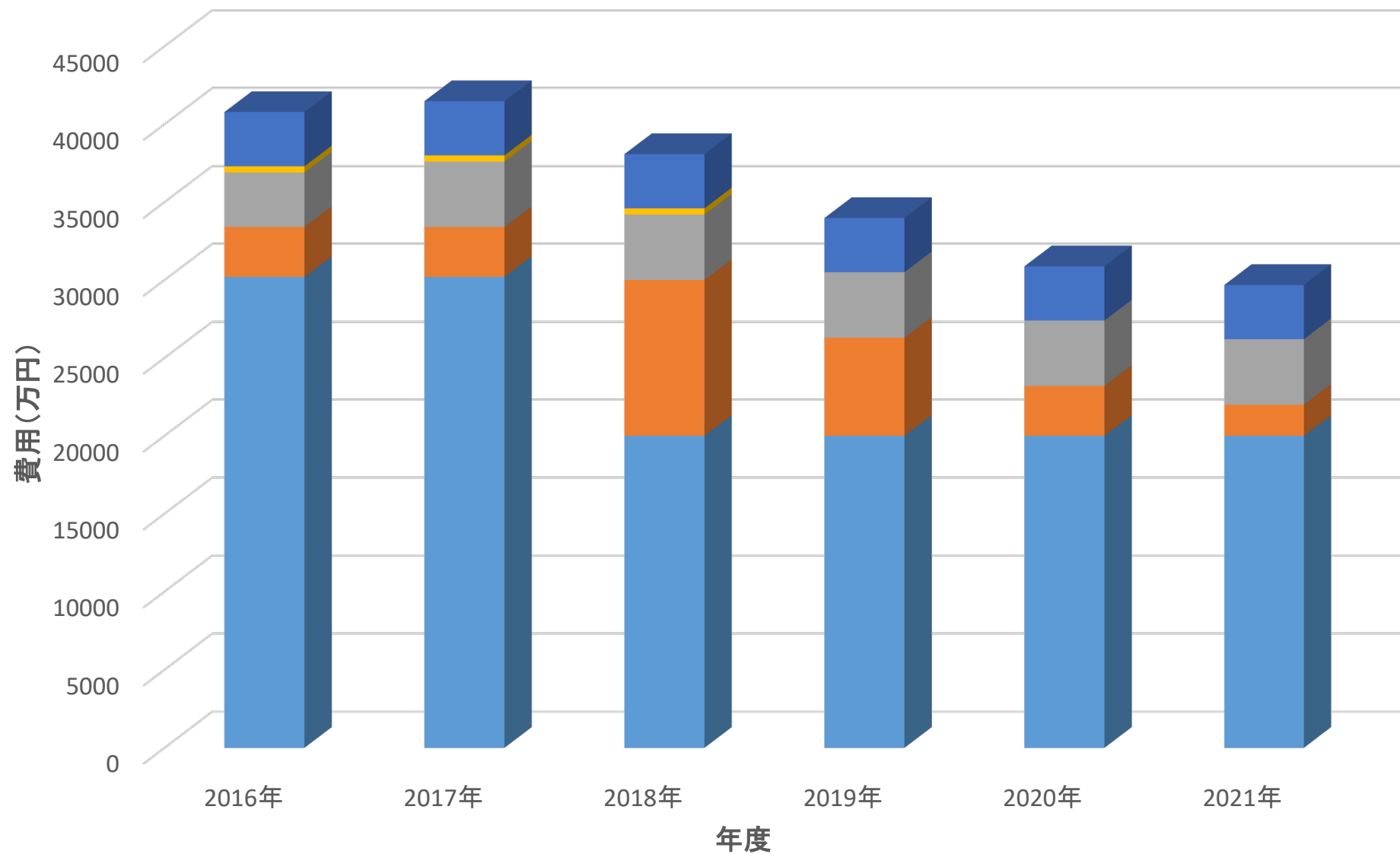
- ✓ レンタル機器の消費電力量 : 50,671 kWh ~ 100万円
  - ✓ 買取機器の消費電力量 : 49,779 kWh ~ 100万円
  - ✓ 空調機関係の消費電力量 : 43,806 kWh ~ 90万円
    - レンタル機器 冷却分 : 45万円
    - 買取機器 冷却分 : 45万円
  - レンタル計算機及びそれを支える空調の電気代は ~145万円/月  
または、1740万円/年
- レンタル契約2億円の9%程度と考えれば良さそう。



# (例)SMOKA運用にかかる概算費用(年額推定)

- レンタル計算機                      ~2420万円 (VizierR、カタログ、ADS等分除外)
  - 買い取り計算機                      400万円
    - 新規機器購入(5年前の機器の更新)
  - 活動費                                 50万円(昨年度までは~300万円)
  - 電気代                                 ~300万円
  - 人件費
    - ✓ 准教授 1名                           1000万円
    - ✓ 研究員 2名                           1000万円
    - ✓ 特任専門員 1名(2月より)         500万円
- 合計 ~5700万円

# ADCデータ系計算機関連予算推移



■ レンタル ■ 買取 ■ 運用人員 ■ 契約人員 ■ 電気代

# まとめ

- アーカイブの運用にはそれなりにお金はかかる。
- 人材も育成しながら、中長期的な視野での戦略を練る必要がある。
- 国立天文台における最近の動きは、我々アーカイブ運用側の自発的なものが少ない。機能改善等は着実に رفتたり、経費の削減については(それなりには)努力しているつもりだが、まだ不足を指摘されることもある。
- アーカイブを使って、科学的成果を最大化するためにはどうすべきかの議論は今まであまりなかった。特にコミュニティーが何を求めているのかを今後どのように吸い上げていくのかは明確化して、仕組みを作っていかないといけない。
- 外的要因は不透明な点も多く、今後も注視が必要。自分たちで何が出来るのかは実はよくわからない。