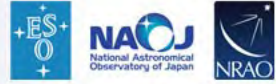




Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins



データの品質管理と利活用の促進

小杉 城治 (国立天文台: NAOJ)

2020/9/14 光天連シンポジウム



Contents:

- ALMA概要
- ALMA望遠鏡の運用について
- ALMAデータの品質保証活動
- ALMAデータの更なる利用促進に向けて
- 最後に



ALMAとは

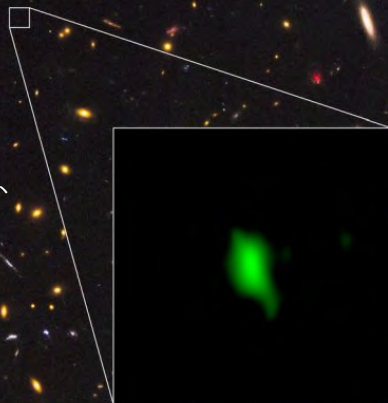
ALMA望遠鏡：人類の新しい眼

- 南米チリ、標高5000mのアタカマ高地に建設、2011年に科学観測を開始。
- 東アジア（日本・台湾・韓国）、北米（アメリカ・カナダ）、欧州南天天文台（16カ国）とチリの国際協力。
- 66台のアンテナ（12m×54台＋7m×12台）を結合してひとつの巨大電波望遠鏡を構成。アンテナの展開範囲は最大16km、山手線大の望遠鏡に匹敵する分解能（0.01秒角＝視力6000）を実現。
- 観測波長はミリ波・サブミリ波（波長 0.35～3.6 mm、周波数 86～950 GHz）。



銀河の誕生と進化

132.8億光年彼方の銀河で検出した、
観測史上最遠の酸素原子輝線



130億光年を超える距離 (赤方偏移 $z > 6$) の宇宙を観測し、
形成直後の銀河の性質や星形成活動を明らかにする。

惑星の形成と進化

うみへび座TW星の原始惑星系円盤

若い星を取り囲むガスと塵でできた原始惑星系円盤を
高解像度観測し、惑星の誕生現場を目撃する。

ALMA望遠鏡が挑む宇宙の謎

生命素材関連物質探査



星形成領域で発見した有機分子
グリコールアルデヒドの想像図

生命の材料となりうるアミノ酸など複雑な有機分
子を、宇宙の様々な場所に探す。





Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins



ALMA望遠鏡の運用について



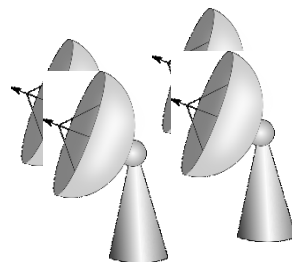
- ALMAは完全なサービス観測
 - 観測提案者(PI)が観測手順書(SB)を準備
 - Observing Tool (OT)に必要な情報を入力することでSBが(半)自動生成される。全ての観測設定情報が含まれる。
 - 時刻、気象条件、望遠鏡配置等によって、適切なSBを選択実行(Schedulerによる選択支援)
 - 観測所がデータの解析と評価
 - PIにデータ配布、その12ヶ月後*に公開

* COVID-19により、最近のデータは占有期間を3ヶ月延ばした

ALMAの運用イメージ



観測提案受付・審査
解像度・感度の指定



観測・データ取得



データ解析
パイプライン
時々マニュアル解析



データ品質保証

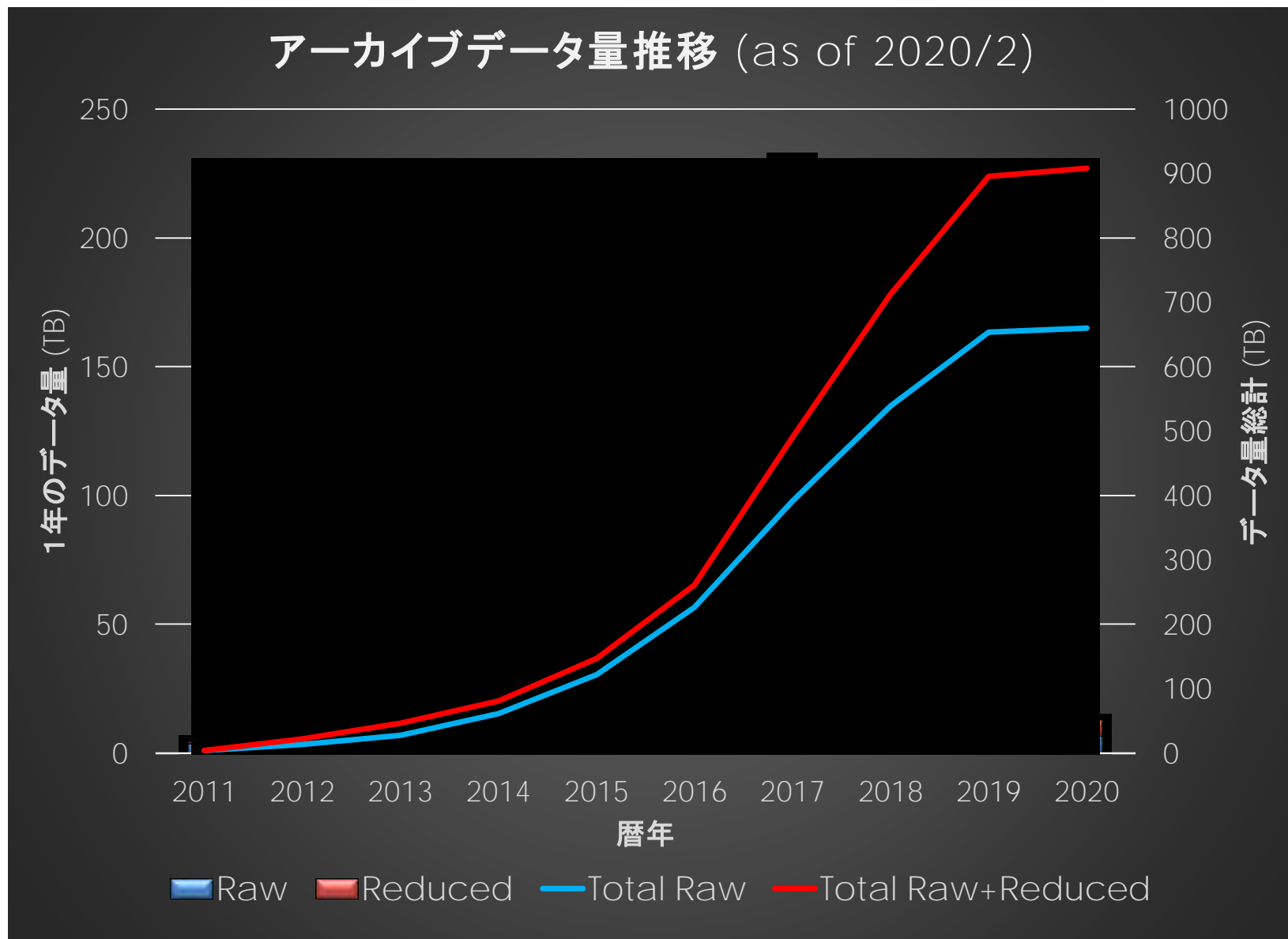


観測提案者への
データ配布・
1年後*に公開

- 最大データレート: 60 MBytes/s
 - 平均データレート: 6 MBytes/s
 - 年間データ量の目安: 200TB/year程度
- ただし、観測データと解析済みデータの合計量
- 最大データレートや年間データ量は数年後にアップグレードされる予定。
- ~700TB/year?



アーカイブデータ量推移 (as of 2020/2)





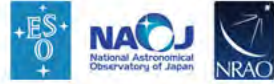
Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins



ALMAデータの品質保証活動



- 観測提案で要求された解像度と感度が達成されたかを観測所が確認
- そのためには、観測所がキャリブレーションを含め、最後までデータ解析をおこなう必要がある
 - ALMAのデータは巨大なため、処理にはそれなりの計算機資源が必要。小さな研究室で解析処理をおこなうのは困難な場合がある（現時点で最大約350GBのデータキューブがPIに配布されている）
 - 解析済みデータをPIに配布することで、PIは論文化に注力できる（ただし、キャリブレーションのみ再利用して、イメージングを処理し直すPIは多い）
 - メリットはPIだけでなく、データ公開後のアーカイブユーザーも享受できる。「[解析済みデータの公開](#)」

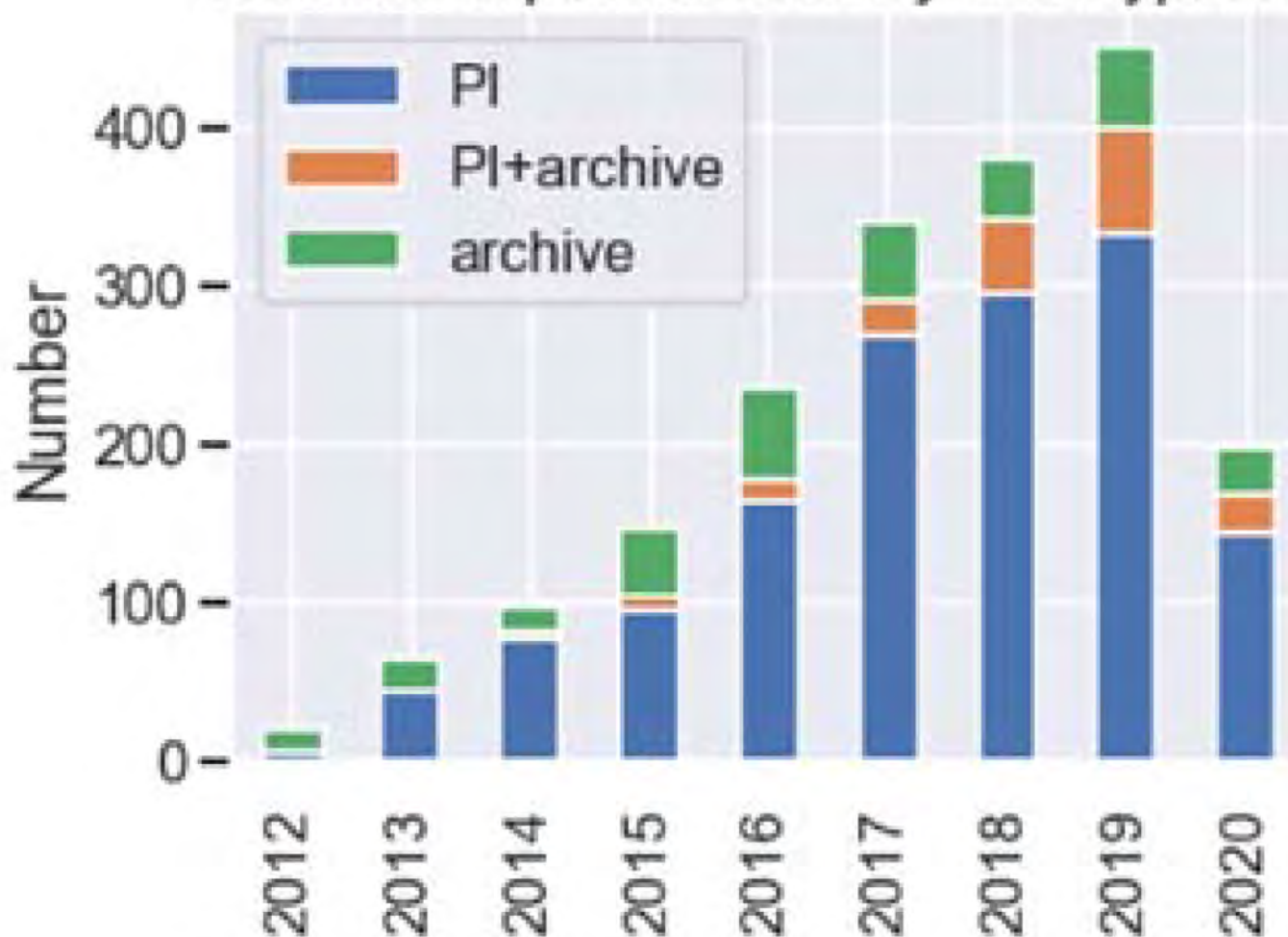


- 新しい(標準)観測モード
- ↓
- 観測データの解析に必要なパイプラインの開発
- ↓
- 観測モードの検証と解析パイプラインの実証確認
(E2E検証)
- ↓
- 次の観測サイクルに新観測モードを提供

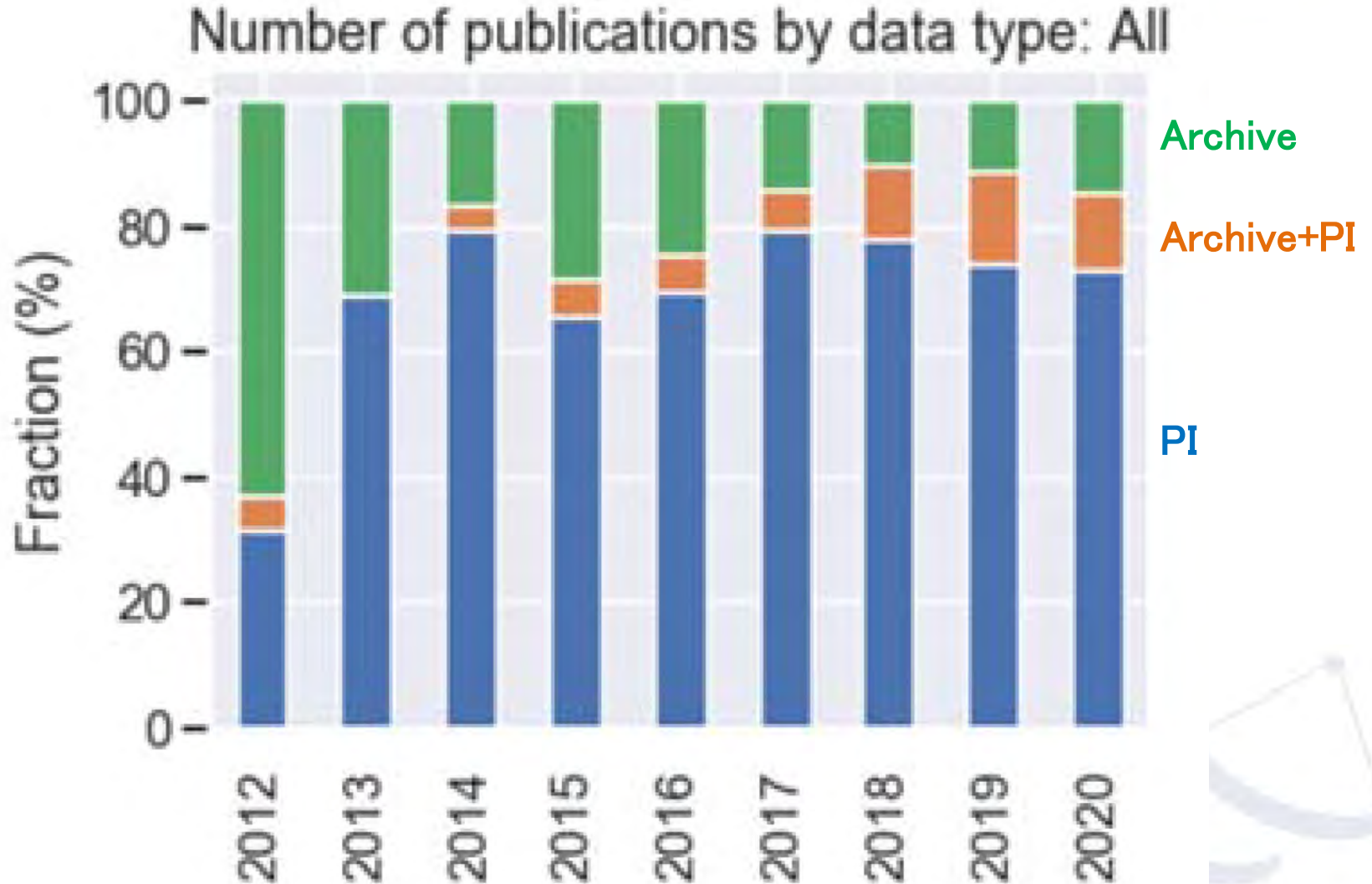
- 解析パイプラインの開発と試験(ソフトウェア)
 - パイプラインは解析ソフトCASAの機能を利用
 - パイプライン初期開発(合計約30人年@construction phase)
 - パイプラインメンテナンスと開発(約5FTE/年:現在)
 - ただし、CASA開発人員は別
- 解析環境の構築と運用(ハードウェア)
 - 日米欧チリで~100台以上の中(大)規模サーバー
- 品質保証活動(人)
 - 日米欧チリで数十人/年規模のサイエンティスト

ALMAデータによる論文生産数の推移

Number of publications by data type: All



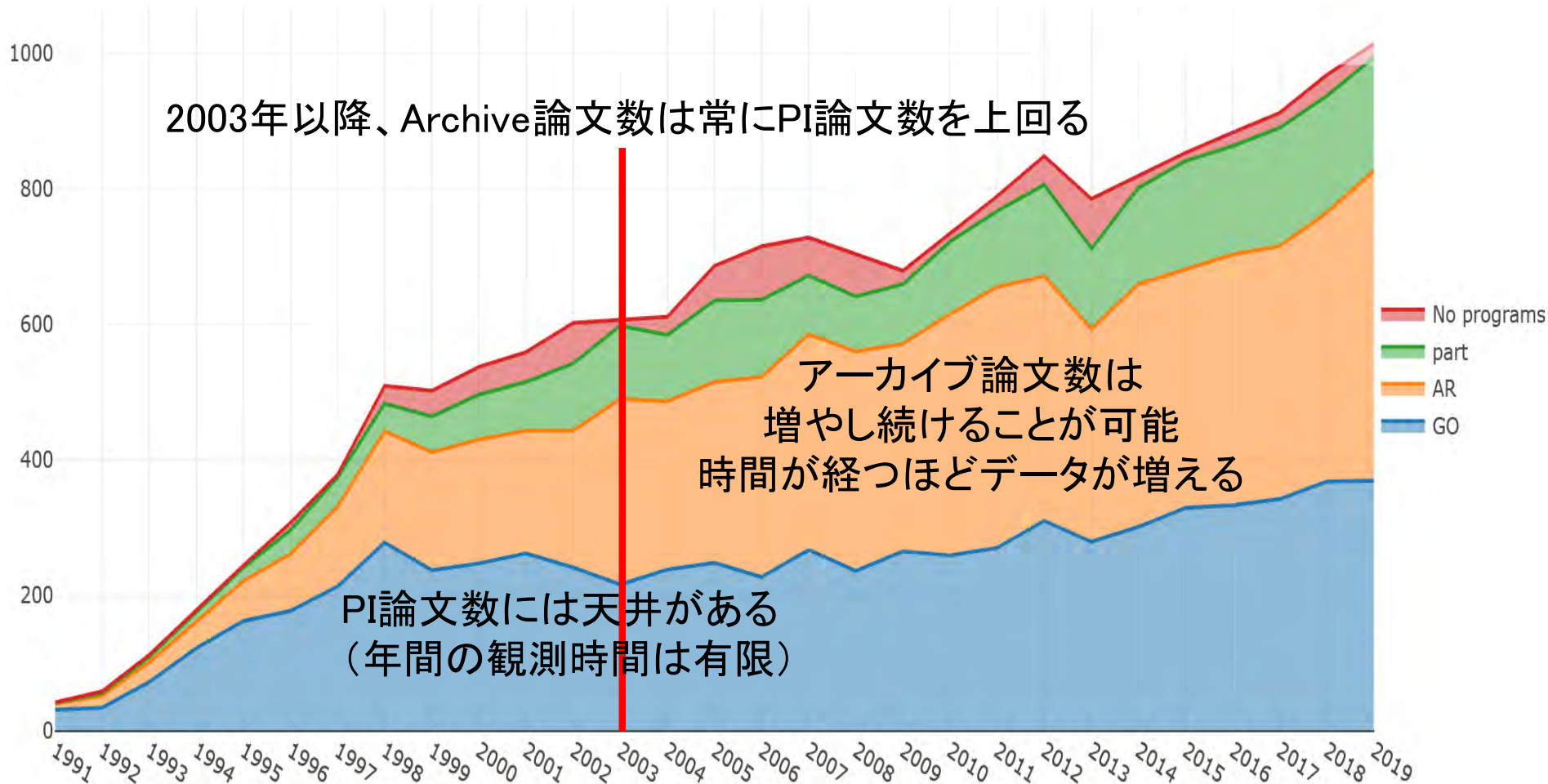
※archiveはScience Verificationデータを含む



※archiveはScience Verificationデータを含む

- ALMAのアーカイブを使った論文数は、2017年以降、増加傾向
 - 最近2年間では全体の25%以上
 - その半分程度がアーカイブデータのみ使った論文
 - まだ増える余地がある(次スライド参照)
-
- 品質保証の年間コストは年間運用費の数%
 - PI論文数に対するアーカイブ論文数を考えると、有意義な投資

参考：HSTのアーカイブ論文数推移



参考: Chandraデータの再利用統計



Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins

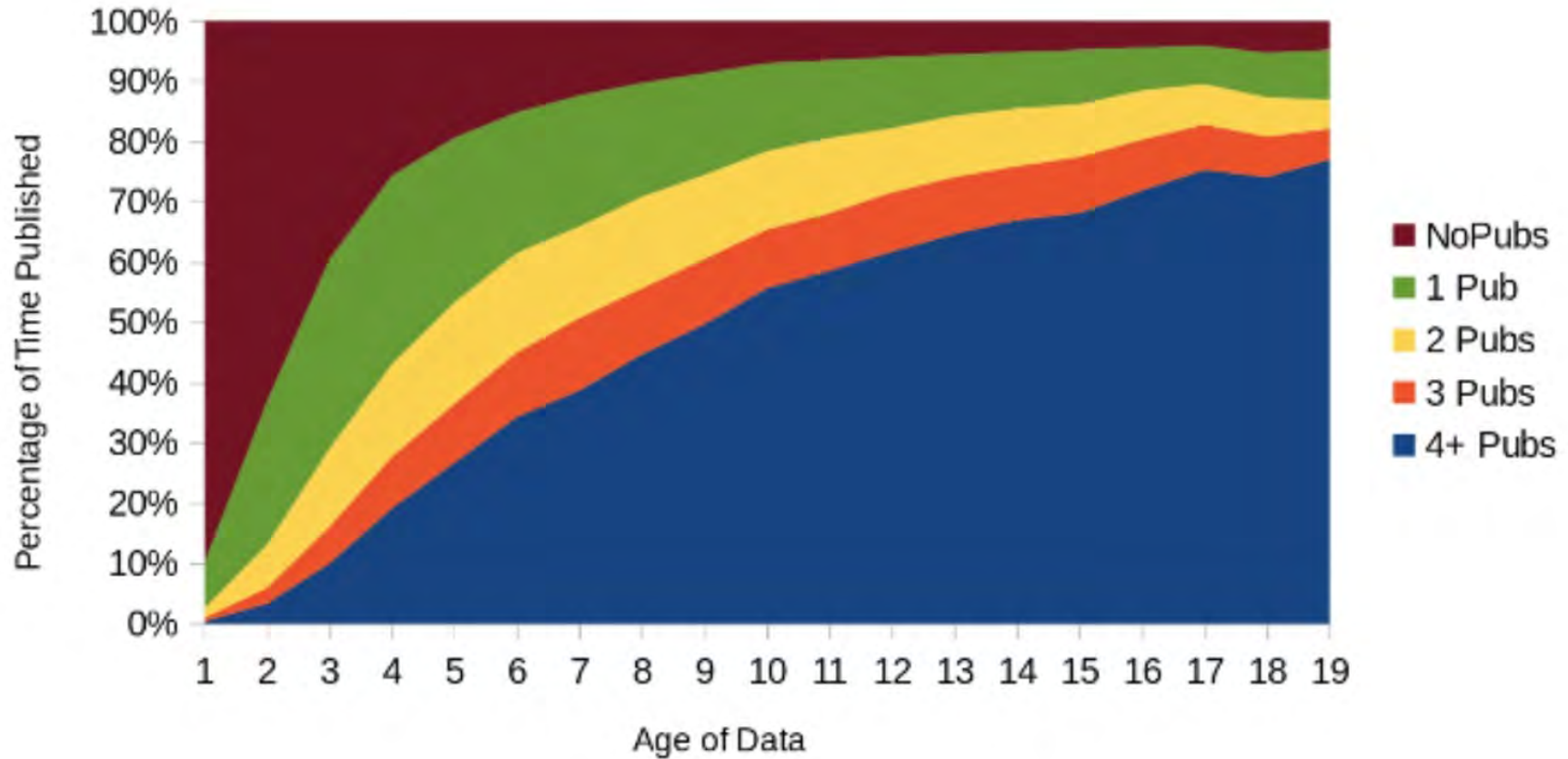


Figure 2. Fraction of the total exposure of Chandra public archival observations published in scientific literature as a function of the time spent in the archive. Different colors show fractions of time published more than once.



Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins



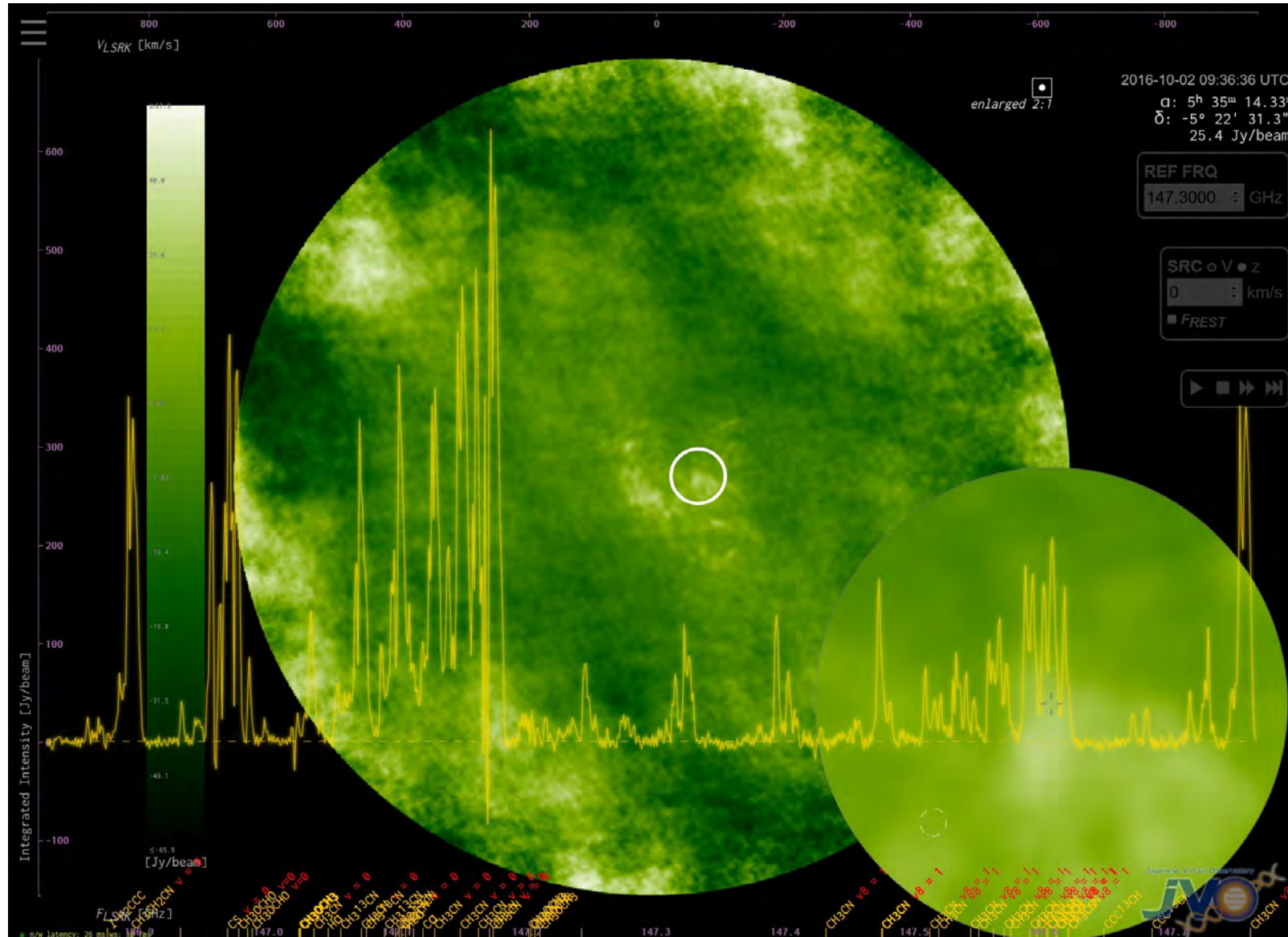
ALMAデータの更なる利用促進に向けて





- データ再利用のハードルを下げる
 - どんな天体がどのような品質でデータに入っているか、ダウンロード前に知りたい。
 - 言い換えると、「ダウンロードして解析するまで使えるデータかわからない」のでは困る。
- JVO Portal / WebQL / Vissageによる可視化
 - <http://jvo.nao.ac.jp/portal/alma.do>





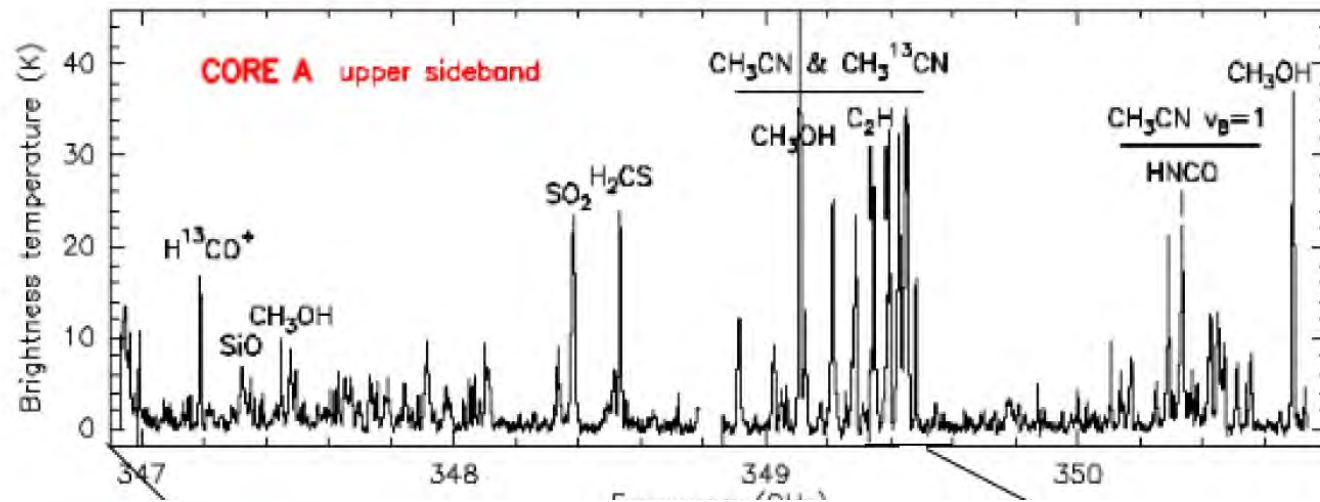


ALMAデータの更なる利用促進②

Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins



- 公開データの品質をより向上させるとともに、利用可能な情報を増やす
 - 最新のパイプラインを使った再イメージング処理
 - 取得された波長範囲を全て利用
- ARI-L (Additional Representative Images for Legacy)
 - Cycle2-4の観測データの再イメージング
 - 既にALMA ArchiveやJVOからアクセス可能



Massardi, M., et al.
ALMA Memo614

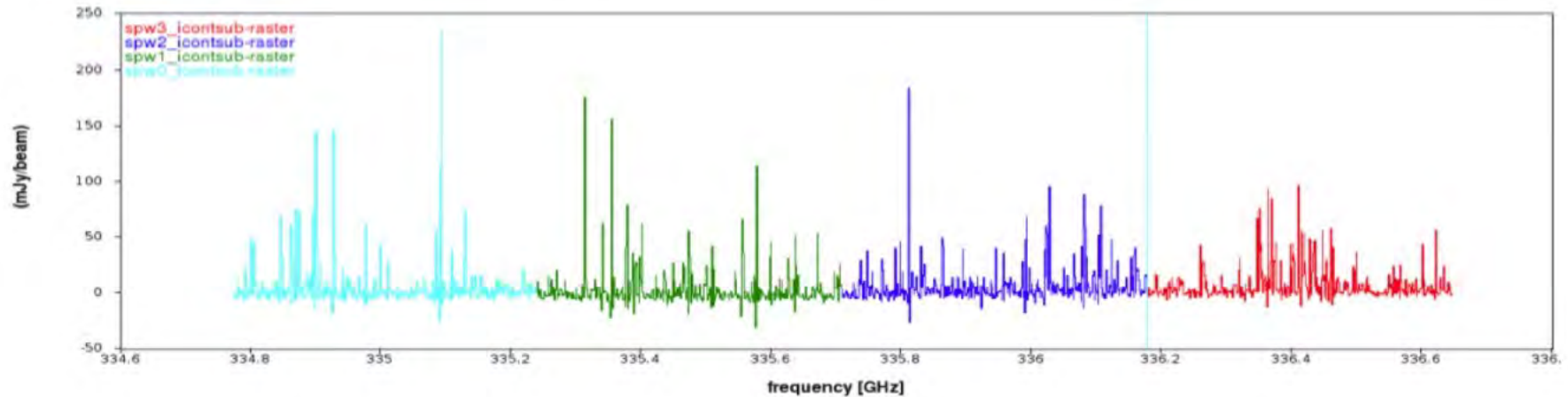


Figure 9: To illustrate the improvement in the archive user's experience provided by the re-imaging efforts, we show, for a target of the same class of objects shown in Fig. 8, the line forest that could be obtained from the re-imaging products in a comparable frequency range.



- 可視化機能の充実化
 - ALMAアーカイブデータの可視化機能も競争の時代に
- CARTA (Cube Analysis and Rendering Tool for Astronomy)
 - coming soon (integrate into ALMA Science Archive)





Menu bar
 File View Layout Help

Widget bar

Dialogue bar

Main window

Status icon

h_m51_b_s05_drz_sci.fits

WCS: (13:29:52.695, 47:11:42.36); Image: (4191, 4746); Value: 5.36169 "ELECTRONS"

Statistics: Region 1 (Active)

Region Active

Statistic	Value
NumPixels	55454 nival(e)
Sum	
Mean	1.8622e+0 "ELECTRONS"
StdDev	2.3461e+0 "ELECTRONS"
Min	5.2067e-1 "ELECTRONS"
Max	3.9851e+1 "ELECTRONS"
RMS	2.9953e+0 "ELECTRONS"

Tool bar

Declination

55

50

45

47:11:40

35

Right ascension

55.5 55.0 54.5 13:29:54.0 53.5 53.0 52.5 52.0

Tab

Render Configuration X Region List X

90% 95% 99% 99.5% 99.9% 99.95% 99.99% 100% Custom

Cursor: 67.02 "ELECTRONS" Value ("ELECTRONS")

Scaling Linear

Color map

Invert color map

Clip Min -0.0104141

Clip Max 1.4575992

X Profile: Cursor X

13:29:55.513:29:55.113:29:54.613:29:54.113:29:53.613:29:53.113:29:52.613:29:52.1

Panel (docked widget)

Data: (4191 px, 5.36e+0) X coordinate

Y Profile: Cursor X

Editing Region 1

Appearance

Color Line Width (px) 2 Dash Length (px) 0

Properties

Region Name

Center (px) 4168.838148 **Dialogue** WCS: (13:29:52.811, 47:11:43.87)

Size (px) 238.201549 233.186780 0.29212' x 0.19431'

Bottom Left (px) 4049.73737 4660.538115 WCS: (13:29:53.395, 47:11:38.04)

Top Right (px) 4287.93892 4893.72489 WCS: (13:29:52.227, 47:11:49.70)

- ALMA望遠鏡の理念

- 電波天文学者以外でも、更に、天文研究者以外でも、ALMAのデータを利用して科学研究ができる

- そのために

- 装置や望遠鏡固有の特性を除去し、データをすぐに使える状態まで整約

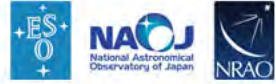
- 観測提案者PIが求めるデータ品質を確認してからデータ配布：品質保証(Quality Assurance: QA)

- そんな高品質な解析済みデータをアーカイブ公開



Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins

Questions?



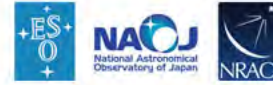
Thank you...

参考：国立天文台の観測データポリシー



Atacama Large Millimeter/submillimeter Array
In search of our Cosmic Origins

天文データ専門委員会 2014/5



前文より

すべてのデータを限られた知識だけで理解することは不可能であるため、データのより完全な理解は、現代のあるいは後世のより広い範囲の英知に委ねられてしかるべき

中略

宇宙をより深く理解するためには、取得したすべての観測データは利用可能な形で保存され、かつ公開されなければならない

1. 国立天文台の観測データは国立天文台に帰属する
2. 国立天文台は、観測データを利用可能なデジタル形式で永続的に保管する
3. 国立天文台は、観測データを利用しやすい形式で公開する
 - データは特定のソフトウェアを用いなくても解析できる水準まで較正処理を進め、できる限りそのまま物理量として扱えるようにした後に公開する。

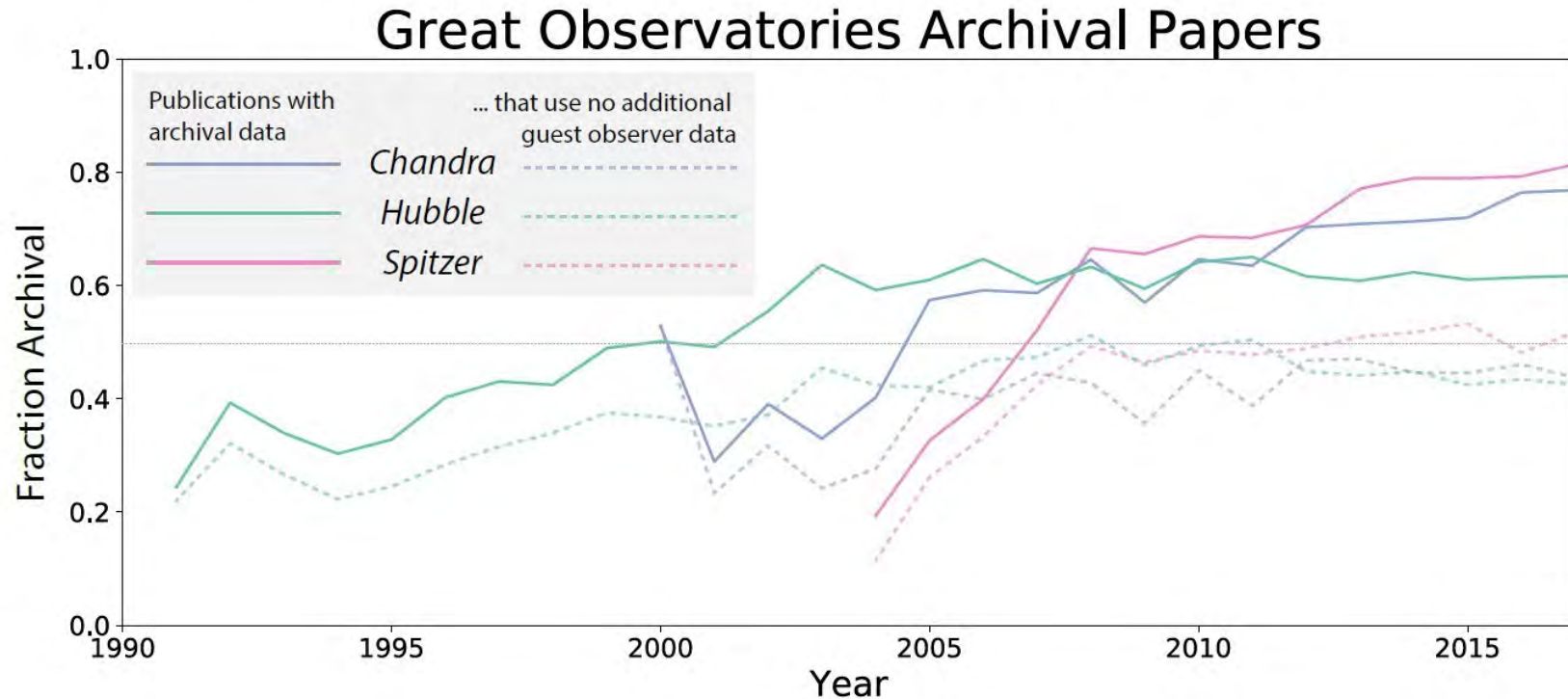


Figure 1: Archival usage of *Spitzer*, *Chandra*, and *Hubble* data. The dashed line represents the fraction of papers that use only archival data from the specified mission. The solid line represents the fraction of papers that use archival data from the specified mission as well as guest observer data.