

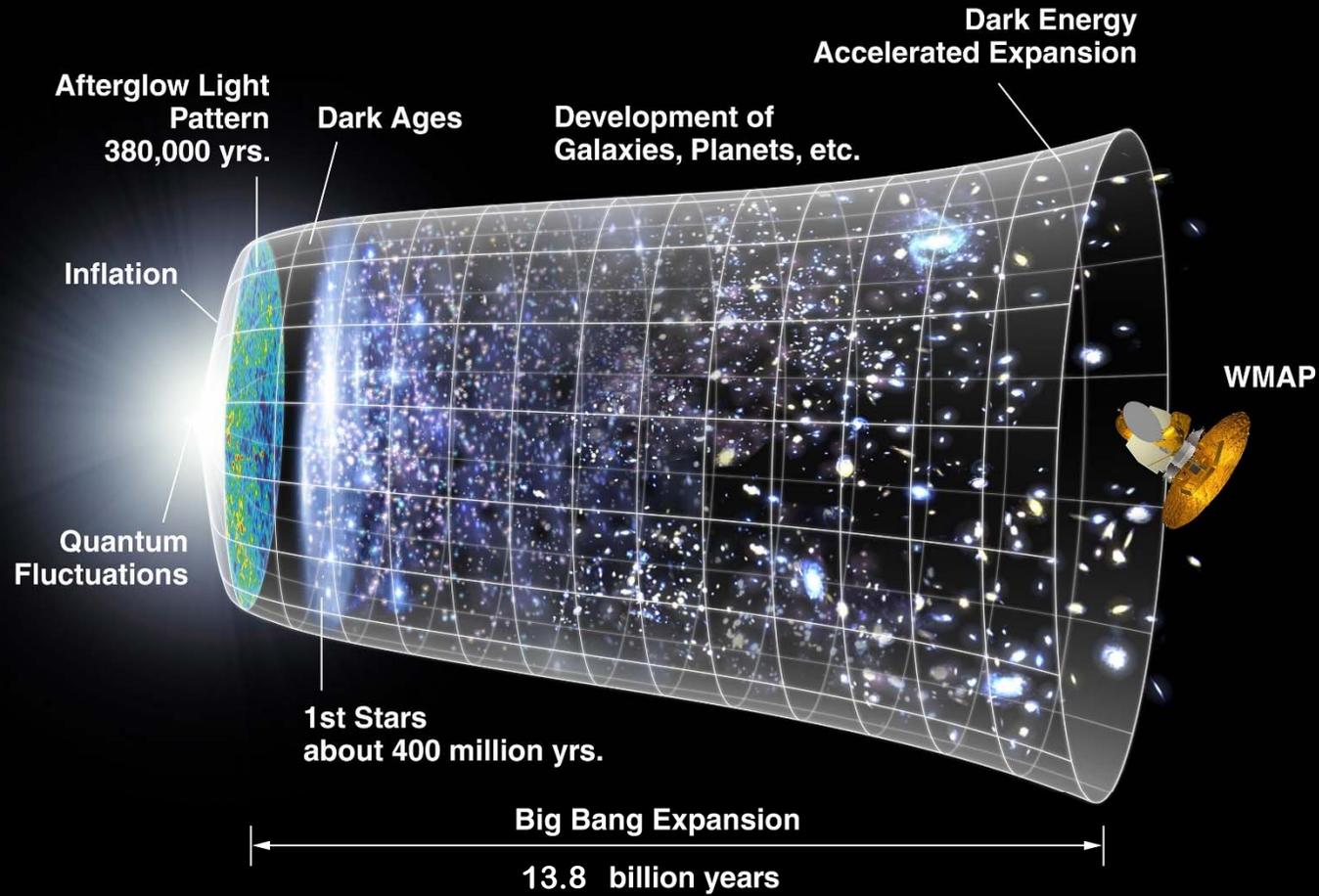
天文学におけるオープンデータ利用： 歴史、機構、成果、そして将来

大石 雅 寿

国立天文台 天文データセンター センター長

masatoshi.ohishi@nao.ac.jp

宇宙の誕生と進化： 銀河、恒星、惑星、生命の誕生と進化の歴史



天文観測今昔物語

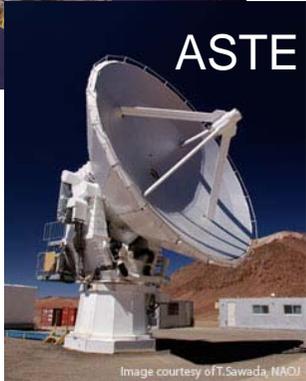
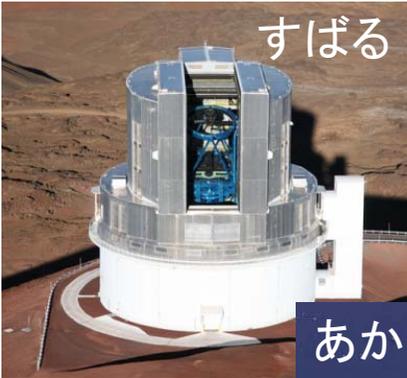
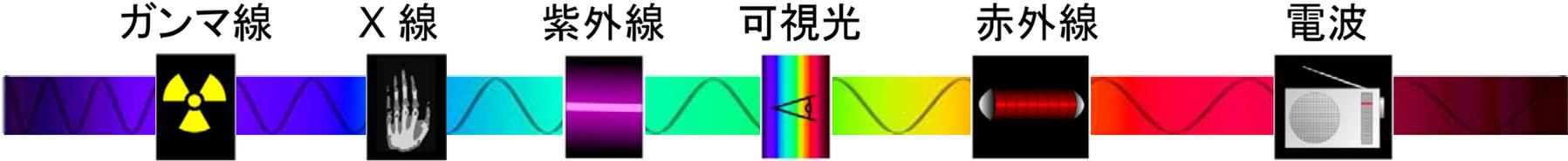


天文学者(一般的イメージ)

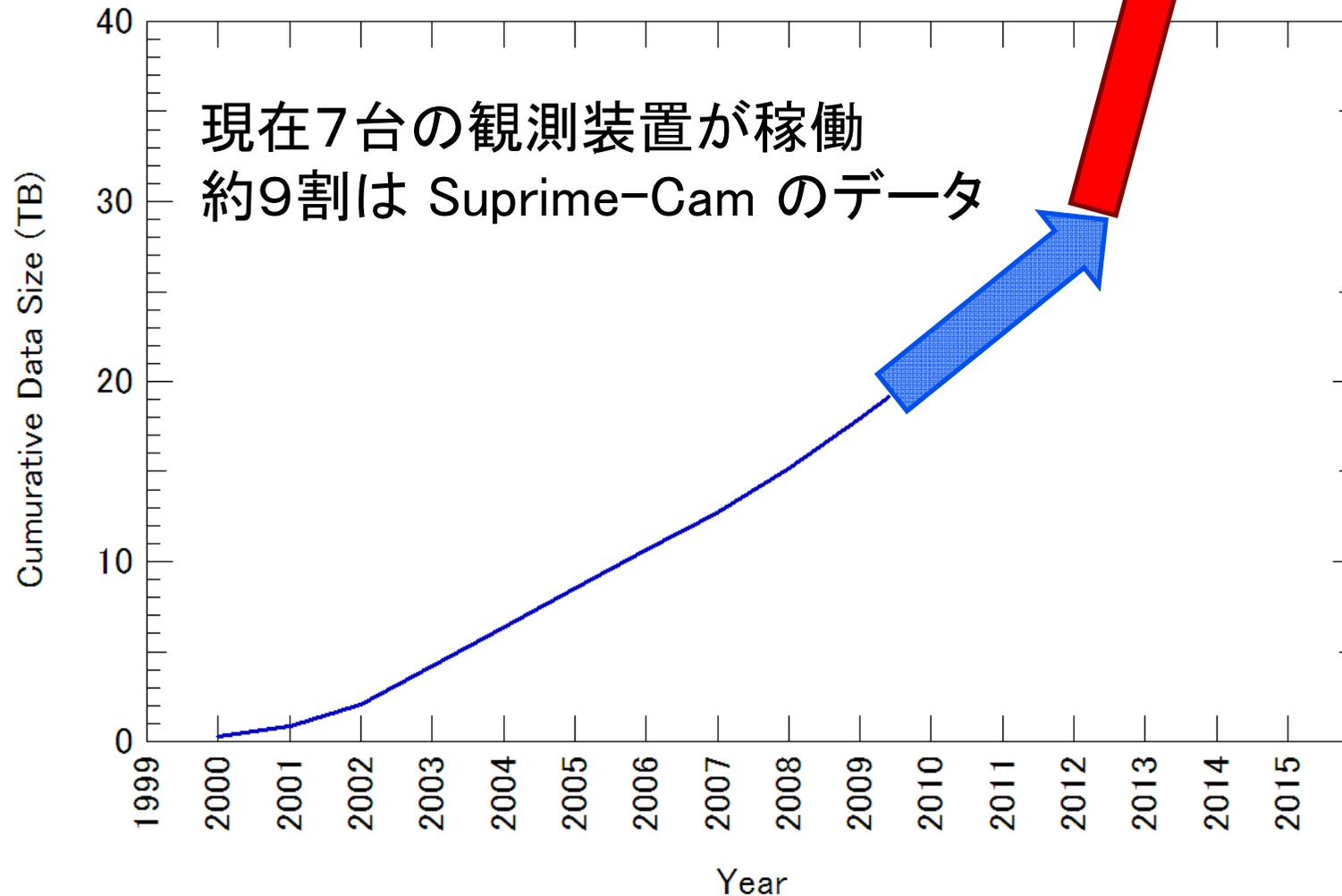


肉眼で観測することはできない。
コンピュータディスプレイ上で確認する。

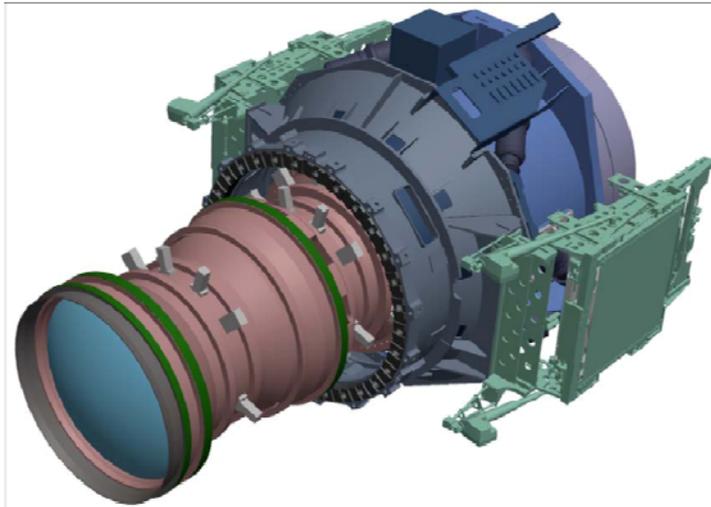
稼働中の主要な望遠鏡 (国内)



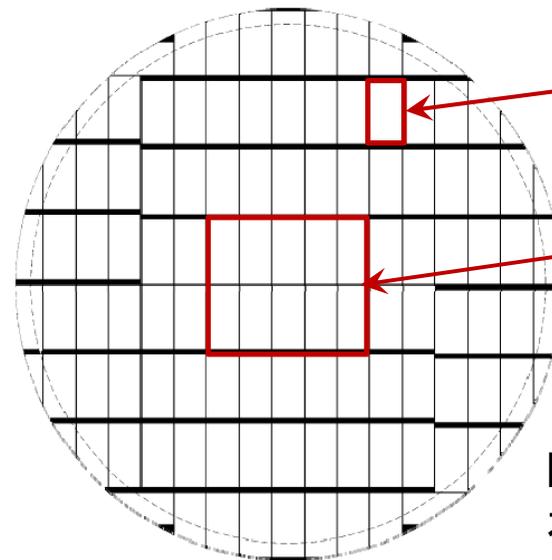
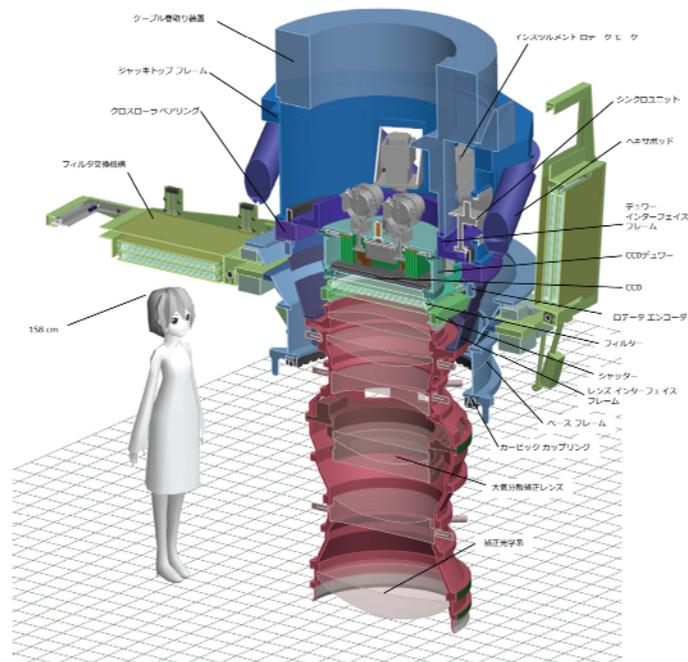
すばる望遠鏡のデータ



Hyper Suprime-Cam (HSC)



- 2k x 4k CCD を 116 枚搭載
- 800メガピクセルの巨大デジタルカメラ
- 一夜あたり 300~500 GB のデータ生成
- 5年間300夜にわたり、最大 2000 平方度の領域を観測予定 → 計 ~150 TB
- 重力レンズ効果を利用した宇宙のダークマター・ダークエネルギーの解明等
- 2014年春観測開始



CCD 一枚

Suprime-Cam

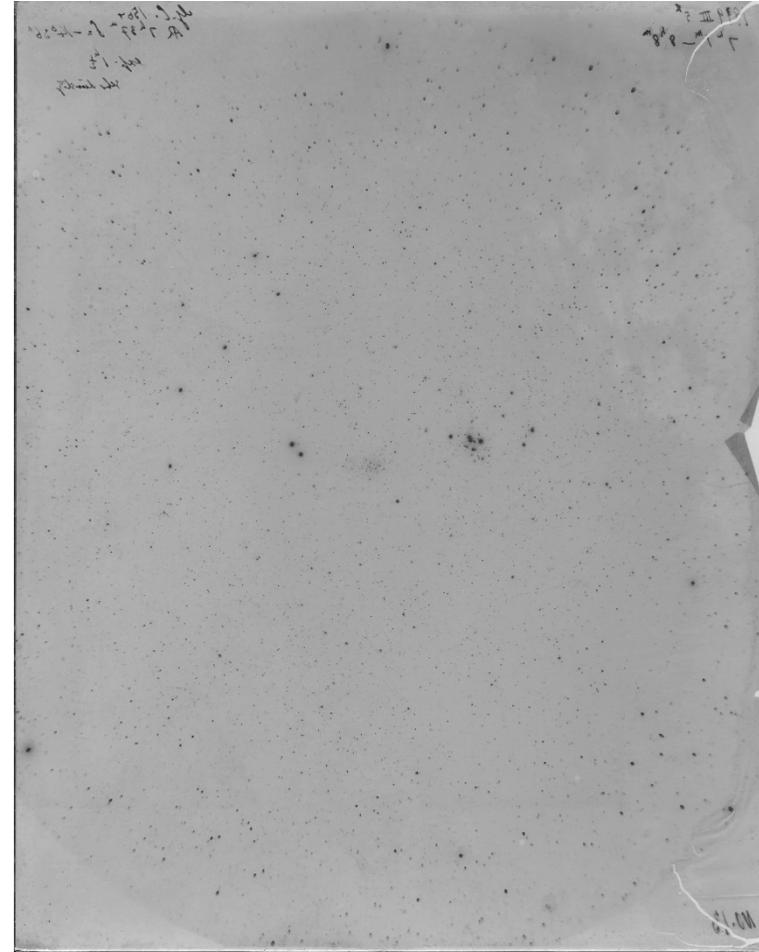
Hyper Suprime-Cam
有効検出領域

天文アーカイブ利用の歴史

天文学アーカイブの必要性(1)

日本最古の写真乾板(1899年3月5日撮影)

- 天文観測データは、観測時点での宇宙のある一角のスナップショットであり、実験により再現することが不可能。このためアーカイブ構築の機運が昔から強かった(写真乾板には約100年の歴史有り)。

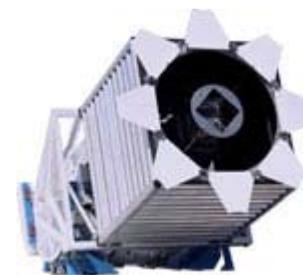


インターネット上の天文研究リソース

- ☐ <http://www.aanda.org/>
- ☐ <http://www.journals.uchicago.edu/ApJ/>
- ☐ <http://www.blackwellpublishing.com/journal.asp?ref=0035-8711&site=1>
- ☐ <http://www.asj.or.jp/pasj/>
- ☐ <http://adsabs.harvard.edu/>
- ☐ <http://www.arxiv.org/>
- ☐ <http://cxc.harvard.edu/>
- ☐ <http://heasarc.gsfc.nasa.gov/>
- ☐ <http://irsa.ipac.caltech.edu/>
- ☐ <http://lambda.gsfc.nasa.gov/>
- ☐ <http://archive.stsci.edu/>
- ☐ <http://nedwww.ipac.caltech.edu/>
- ☐ <http://nssdc.gsfc.nasa.gov/>
- ☐ <http://www.spitzer.caltech.edu/>
- ☐ <http://cdsweb.u-strasbg.fr/>
- ☐ <http://cfa-www.harvard.edu/iauc/SearchIAUC.html>
- ☐ <http://www1.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cadc/>
- ☐ <http://skyview.gsfc.nasa.gov/>
- ☐ <http://archive.eso.org/>
- ☐ <http://dbc.nao.ac.jp/>
- ☐ <http://idlastro.gsfc.nasa.gov/>



- <http://nrodb.nro.nao.ac.jp/>
- <http://www.darts.isas.jaxa.jp/>
- <http://www.sdss.org/>
- <http://www.ipac.caltech.edu/2mass/>
- <http://www.ukidss.org/>
- <http://www.astro-wise.org/>
- <http://terapix.iap.fr/>
- <http://www.roe.ac.uk/ifa/wfau/>
- <http://www.jach.hawaii.edu/UKIRT/>
- <http://www.cfht.hawaii.edu/Science/CFHTLS/>
- <http://swire.ipac.caltech.edu/swire/swire.html>
- http://www.oamp.fr/virmos/virmos_vvds.htm
- <http://deep.ucolick.org/>
- <http://www.eso.org/science/eis/>
- <http://www.galex.caltech.edu/>
- <http://www.stsci.edu/science/goods/>
- <http://www.ast.cam.ac.uk/~wfcsur/>
- <http://www.noao.edu/noao/noaodeep/>
- <http://www.esa.int/SPECIALS/ESAC/index.html>
- <http://www.eso.org/public/astronomy/archive.html>

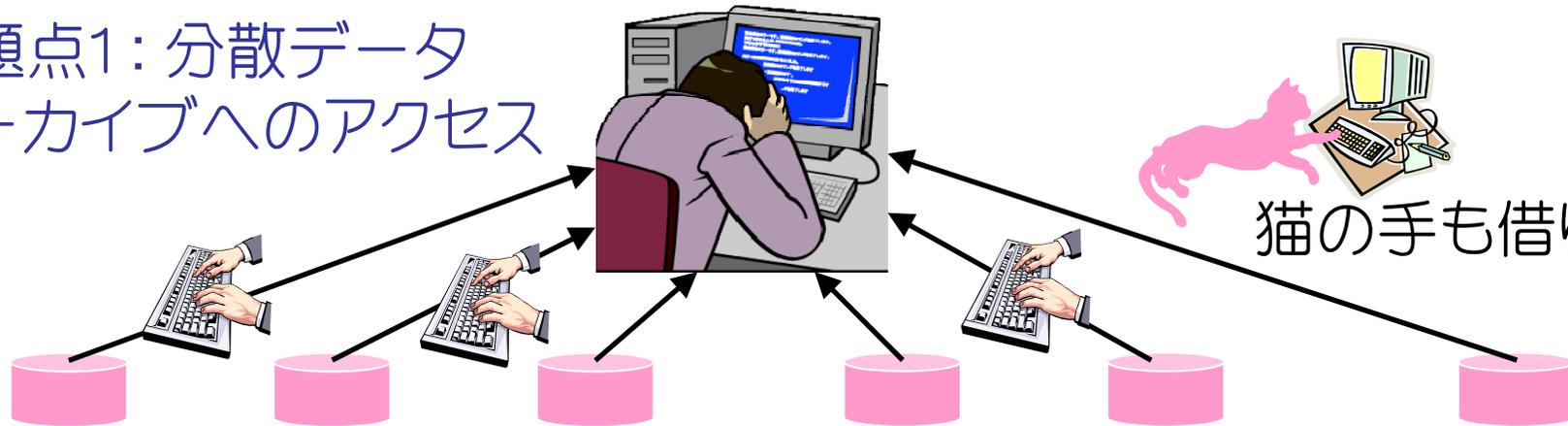


文献 & カタログデータベース

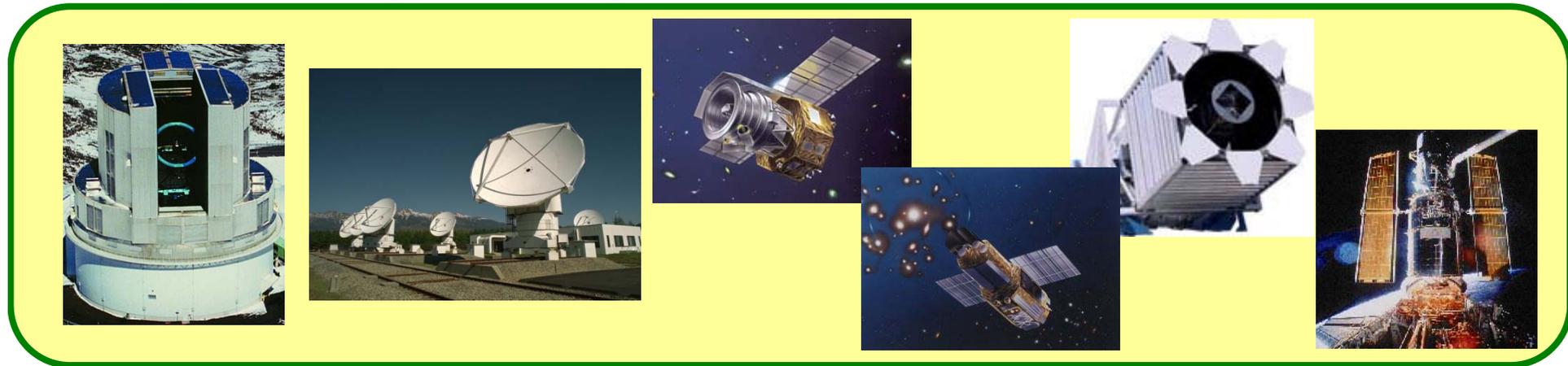
- 文献データベース — ADS (NASA)
 - 主要な天文系論文をオンライン公開、> 10M records
- データやカタログのみの天文論文も刊行
- CDS (Centre de Données Astronomiques de Strasbourg)
 - 天体カタログの頒布(1972年～)
 - SIMBAD: 天体のメタ情報DB、700万天体
 - VizieR: 論文に掲載された表などをonline公開

天文アーカイブ利用に関わる 新たな課題

問題点1: 分散データ
アーカイブへのアクセス

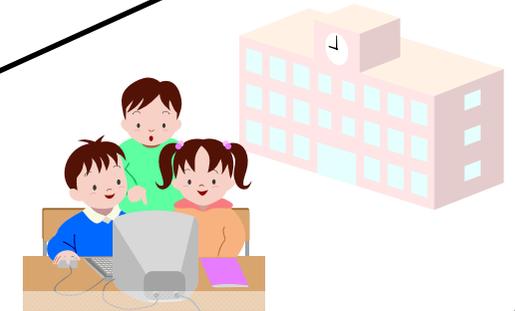


猫の手も借りたい



データベースアクセスインターフェースの共通化

データ
取得を
自動化



教育用教材としても利用可³

Virtual Observatory (VO)

デジタル化された天文データを**観測**し、そのデータを解析・処理することにより天文学的知見を生み出す**抽象化された観測装置**

いつでも、どこからでも、天候などに左右されずに観測することができる研究インフラ

大量データの統計処理を容易に実行することによって、天文研究の質的転換を目指す

International Virtual Observatory Alliance

- 世界各国の 21の VO プロジェクトが参加
- 天文データの共有をより効率的に行うための標準仕様策定団体

<http://www.ivoa.net/>



- 策定された主な仕様 (全部で30以上の仕様)

IVOA Registry Interfaces – データサービスの公開方法

Resource Metadata for the Virtual Observatory – データサービスのメタデータ仕様

VOTable Format Specification – 検索結果等のデータフォーマット

Simple Cone Search

Simple Image Access

Simple Spectral Access

Table Access Protocol

} データ検索のインターフェイス定義

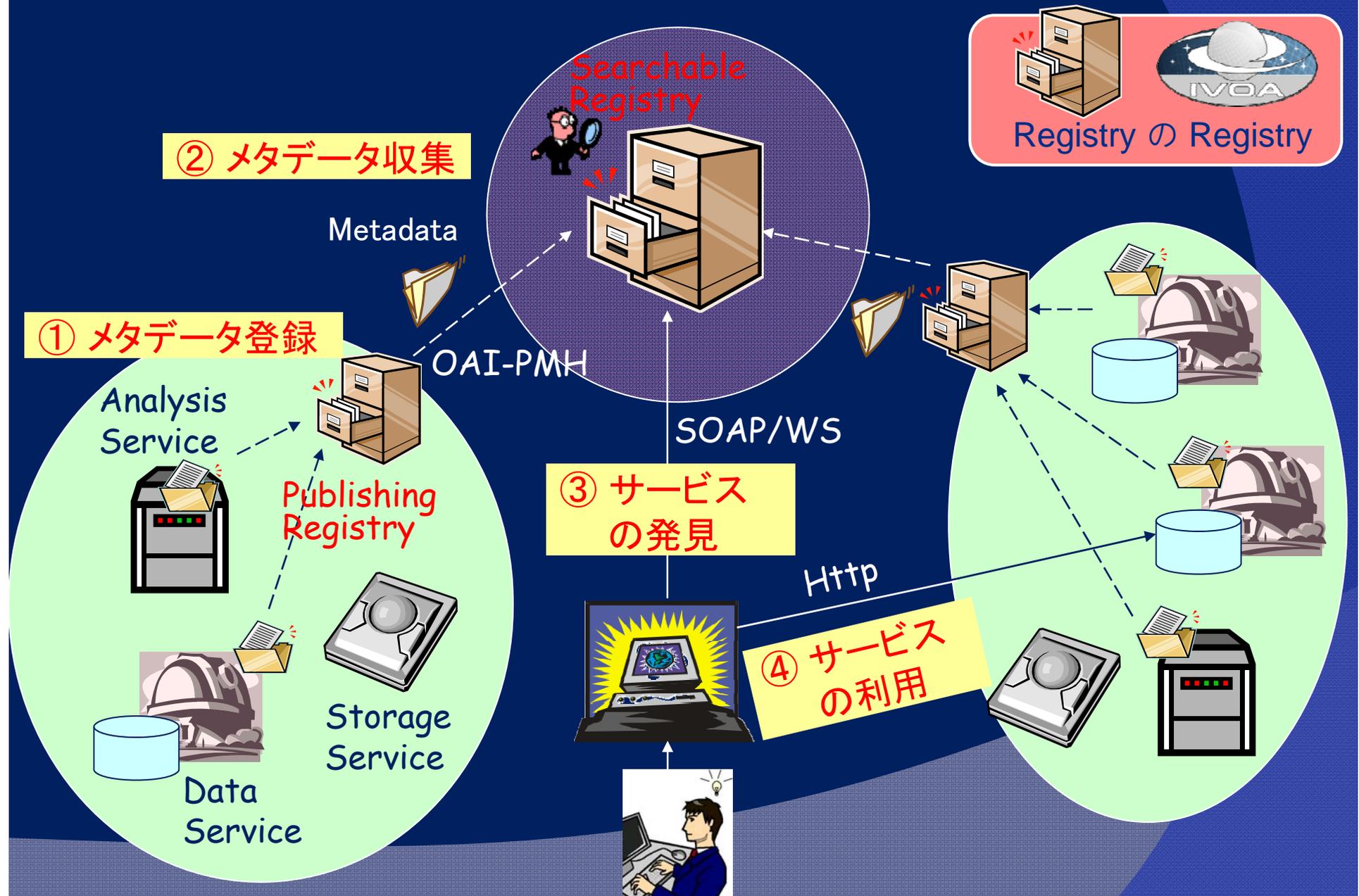
IVOA Astronomical Data Query Language – 検索言語仕様

VOSpace service specification – 分散ファイルシステムを実現する仕様

Data Model for Astronomical DataSet Characterisation – データモデル

Simple Application Messaging Protocol – アプリケーション間連携の仕様

データサービスの公開と利用の仕組み



標準策定作業

- **標準化は非常に有効**
 - Access protocols, data format, etc.
 - Interoperability → wider dissemination and application
 - Endorsement by the IAU (VO WG)
- **苦痛を伴うプロセス**
 - Philosophy, intention, life time of project,,
 - Compromise, patience
 - Establishment of relationship : respect to each other
 - Coffee/tea breaks and lunch/dinner talks are crucial

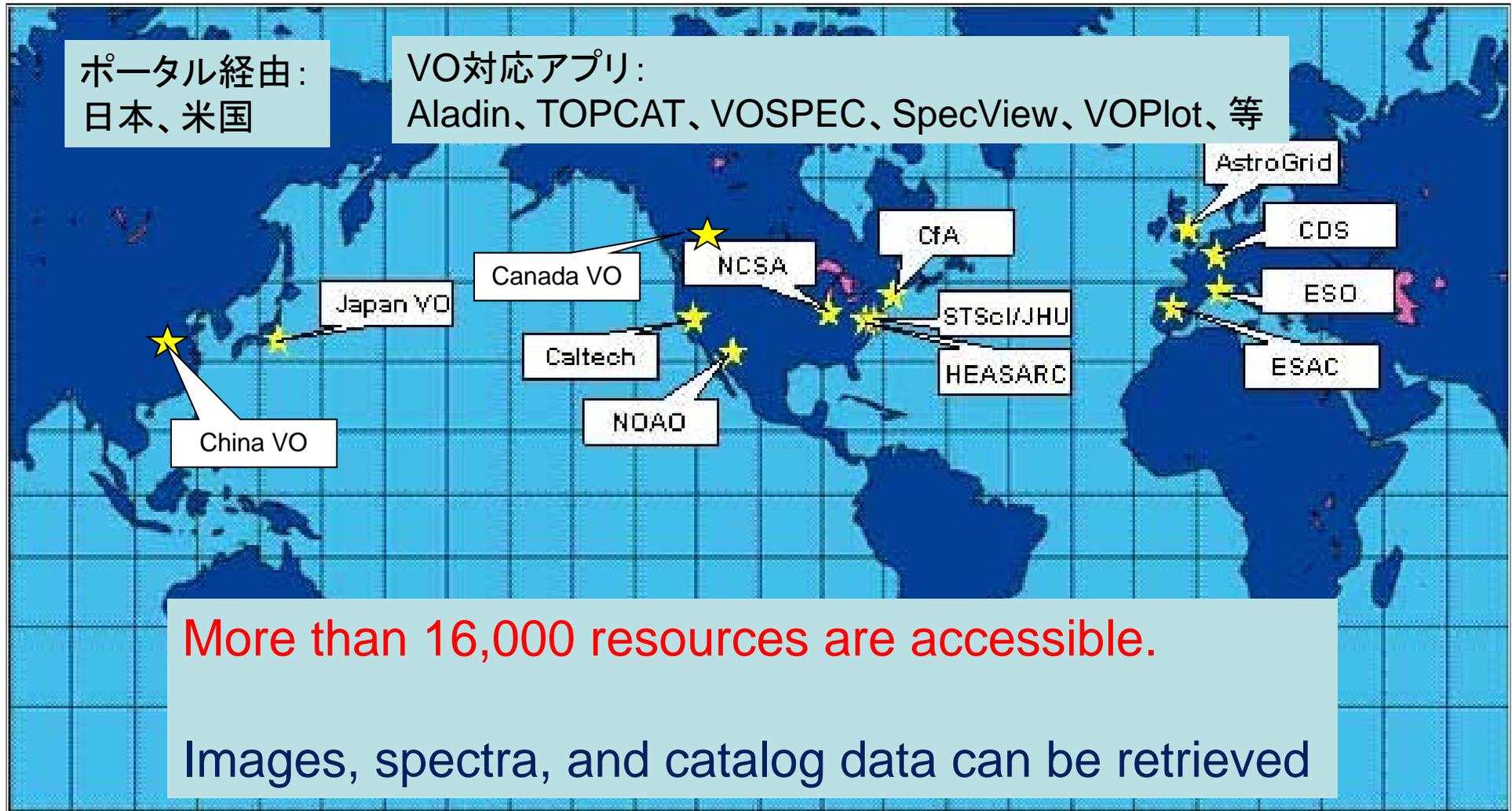
IVOA 相互運用会議



2010年12月 奈良にて

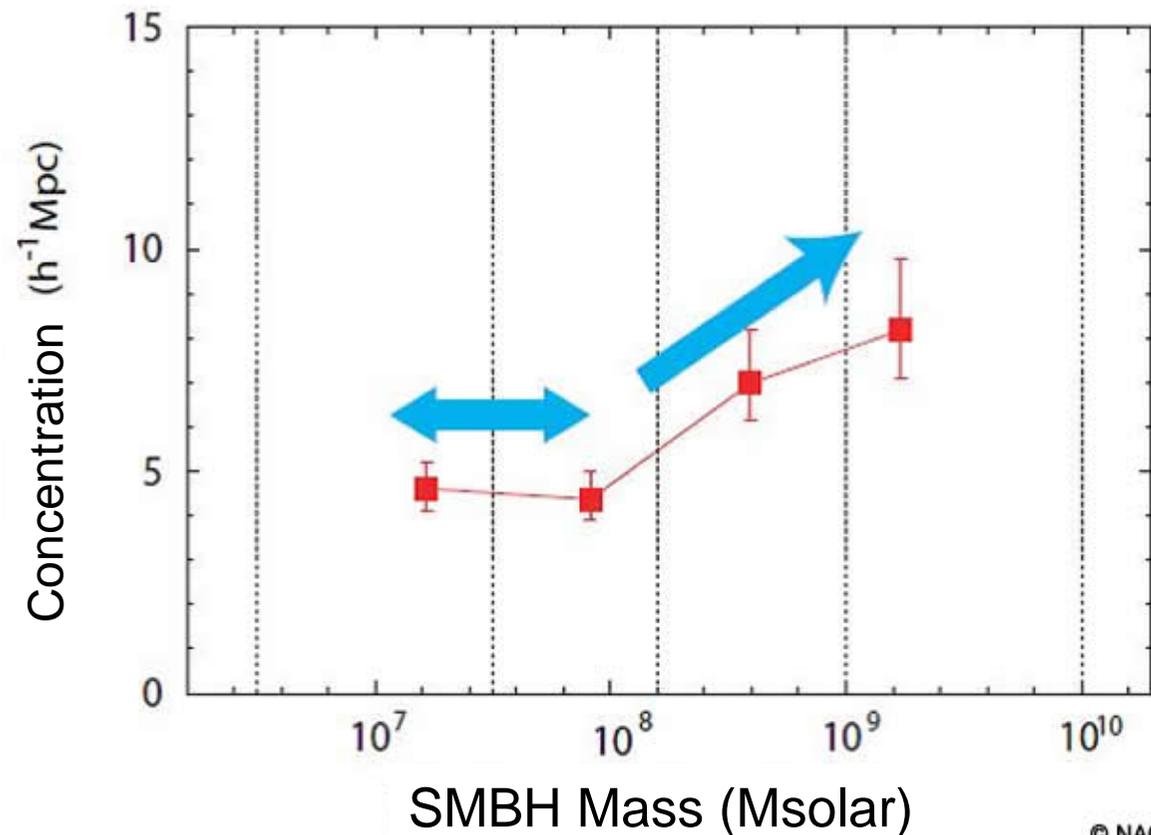
- 2003年以来、毎年2回開催
- 標準化に向けた集中議論 (キツイ!)
- 技術に強い若手が中心に活躍
- ネットを介した協働に必須な人脈形成 (Layer 0)

VOにより連携する世界の天文台



Science output by JVO

- Concentration of galaxies around SMBH
 - 10,000+ SMBH & 7×10^7 galaxies collected by JVO
- Komiya et al., ApJ, 775, 43 (2013)



VOを使って生み出された
査読論文
1073 (ADS調べ)

2002年1月～2015年6月

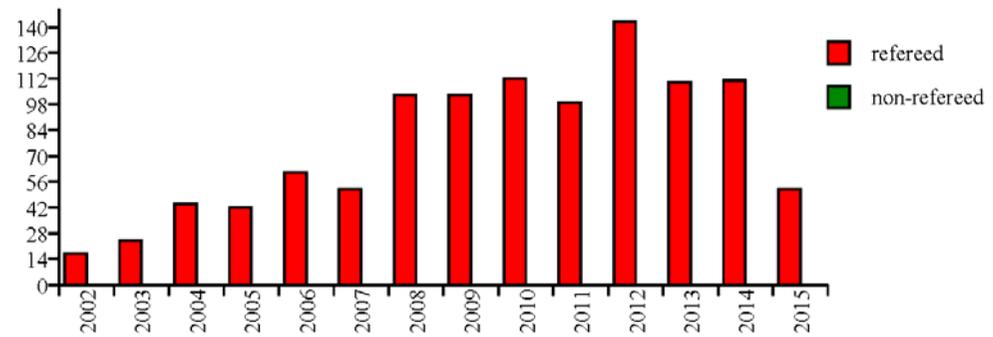
2015年6月11日

READ10 index

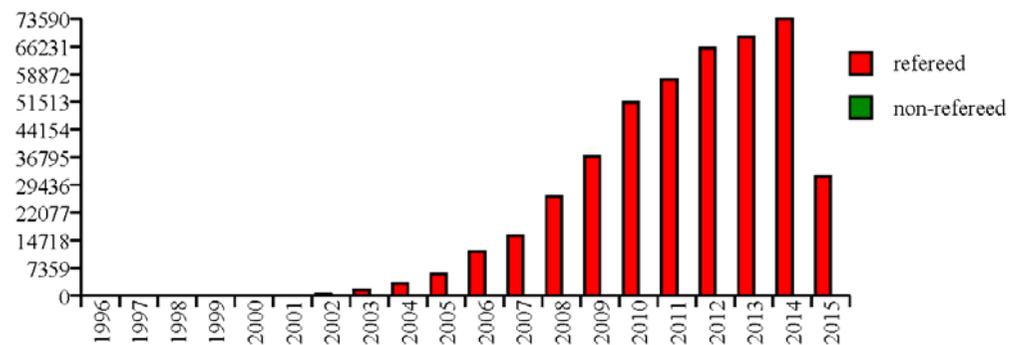
12989

12989

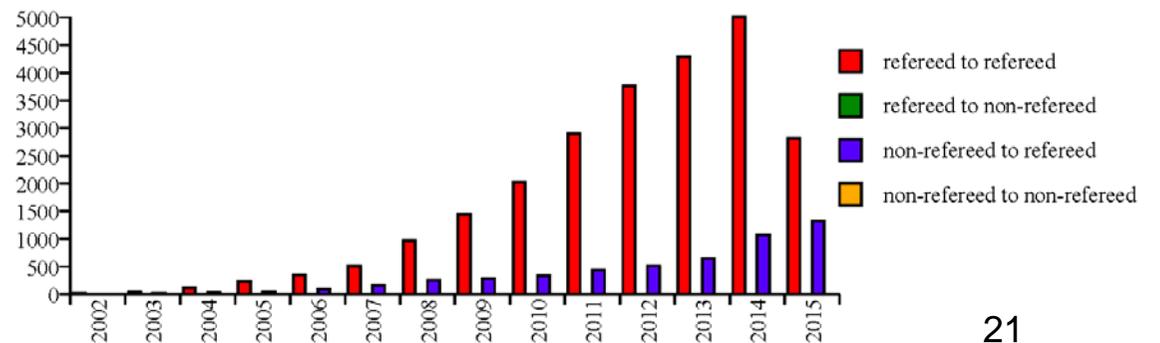
publication histogram



reads histogram



citation histogram



不思議なことに海外からのアクセスが多い

March, 2015

Top 10 of 60 Total Countries							
#	Hits		Files		KBytes		Country
1	410853	84.82%	374463	85.19%	8142544	11.58%	Japan
2	15492	3.20%	13863	3.15%	4885873	6.95%	Unresolved/Unknown
3	12534	2.59%	11250	2.56%	6096353	8.67%	Commercial (com)
4	10327	2.13%	9179	2.09%	55650	0.08%	Spain
5	10324	2.13%	10323	2.35%	1810138	2.57%	US Government (gov)
6	6678	1.38%	6095	1.39%	60739	0.09%	France
7	4528	0.93%	4190	0.95%	48760	0.07%	International (int)
8	3885	0.80%	2826	0.64%	2328270	3.31%	Network (net)
9	3004	0.62%	2428	0.55%	93367	0.13%	Germany
10	2313	0.48%	2291	0.52%	35676807	50.74%	Chile

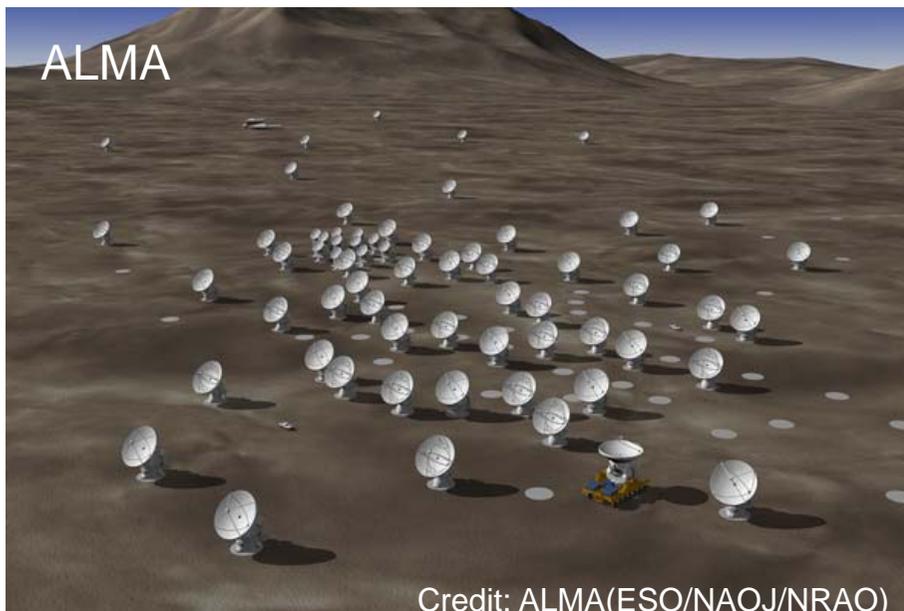
欧米と日本の違いはどこから？

- 欧米でのアーカイブ利用は、日本の数倍
→ Publish or Perishのため？
- 科学研究は、データから如何にして新知見を引き出すかにかかっているはずだが・・・
→ 日本では「もの作り」だけで息切れ？
- 日本人は「もの作り」を重視
→ ソフトウェアやデータはおまけ？

日本人の慣習、文化、心情を踏まえた普及対策が必須

将来に向けて

大規模データを生成する望遠鏡計画



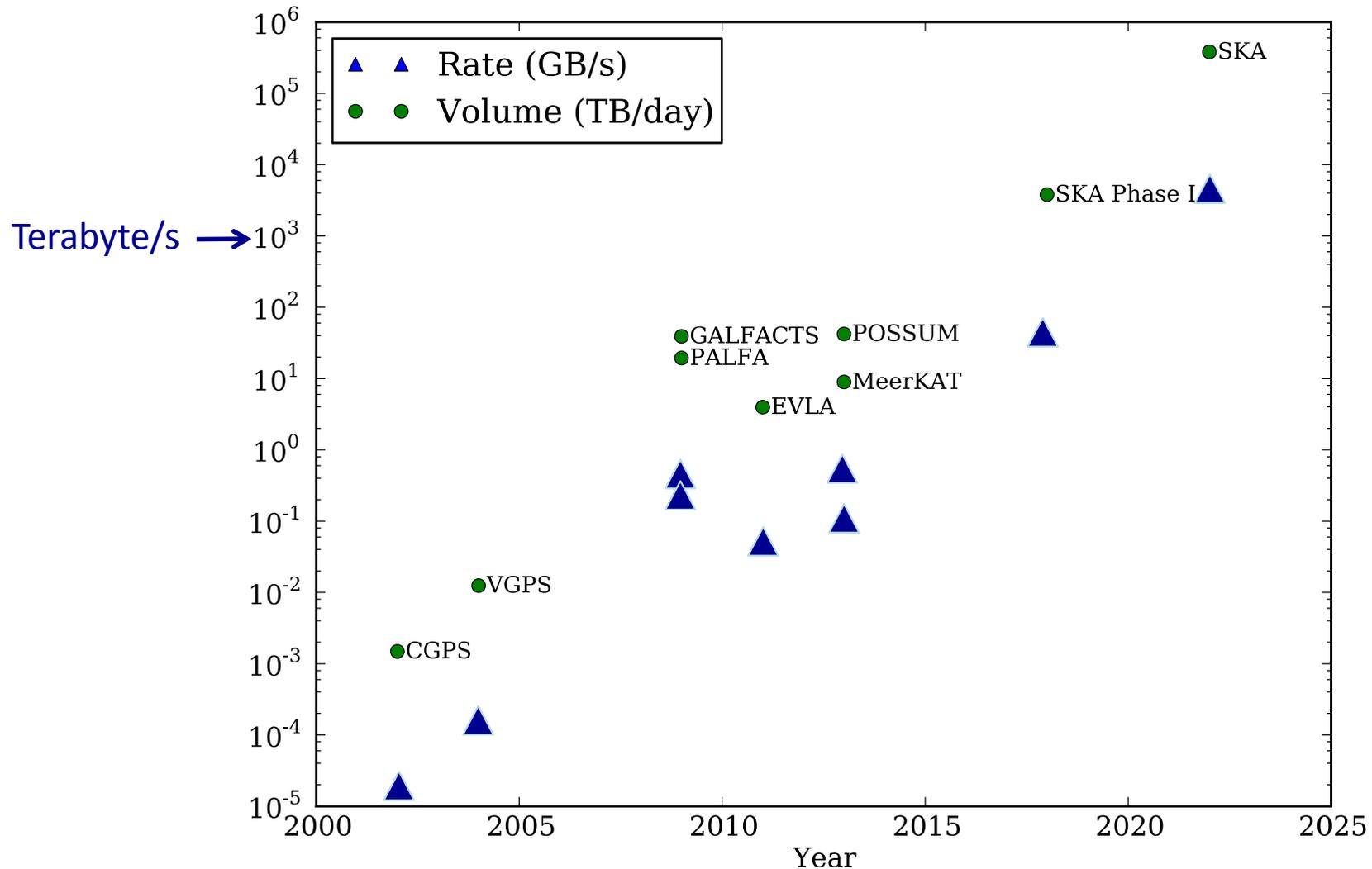
- 66台のアンテナ @アタカマ砂漠 標高 5000m
- 主として日・米・欧の諸国が協力して開発
- ハッブル宇宙望遠鏡の約10倍の空間分解能 (0.01秒角)
- 宇宙初期の生まれたての銀河、恒星や惑星の誕生過程、物質進化の探求から生命の起源等を探る
- 年間 200 TB超のデータが生成される。
- 2011年観測開始

- 8.4m 光赤外望遠鏡 (米国 2018?~)
- 3.2 ギガピクセルの巨大カメラ(視野は約10平方度)
- 3晩で観測可能領域をすべて撮像。同じ領域を3晩毎にくり返し観測 → 宇宙のムービー
- ダークエネルギーの解明、太陽系新天体の探査、突発天体の探査等を行う。
- 一晩で15テラバイト、1年で6.8ペタバイト
- Google が解析システムの開発に参加

LSST



Survey Raw Data Output Rates



IAU GA Beijing August, 2012



- Scientific Impact of Past and On-Going Large-Scale Observations and Surveys to Astronomy
- Current Status and Challenges of Future Large-Scale Observations and Surveys (1) Near- and Mid-future projects, (2) Far-future projects
- Data Management and Data Access: Past, Present and Future
- Advanced Data Analysis in the Data Intensive Astronomy Era
- Synergy of Data Intensive



High level Data Analysis

- 巨大データからの”新ルールや知見“の発見
 - Needles in haystacks – Higgs 粒子
 - Haystacks: Dark matter, Dark energy
- 概略解の探求、分類、、
新しいアルゴリズム – Data Mining (KDD)
- 統計学やコンピュータ科学との協同が必須

World Data System



- ICSU (国際科学会議)
Strategic Coord.
Comm. on Info. &
Data
- 地球物理, 農業, 環境,
天文等の科学データを世界規模で共有 (途上国に提供) するためのフレームワーク
- 天文VOが参照例の一つに

<http://www.icsu-wds.org/>



Summary

- 天文分野におけるデータ共有機構として、国際的な標準に基づいたヴァーチャル天文台が稼働。
- VO利用は欧米が圧倒的に多い。日本がまだ少ないのは文化的な違いによるのか？
- 多様な分野におけるオープンなデータ利用やサイエンスを推進する政策が取りまとめられつつある。 Be ready !