

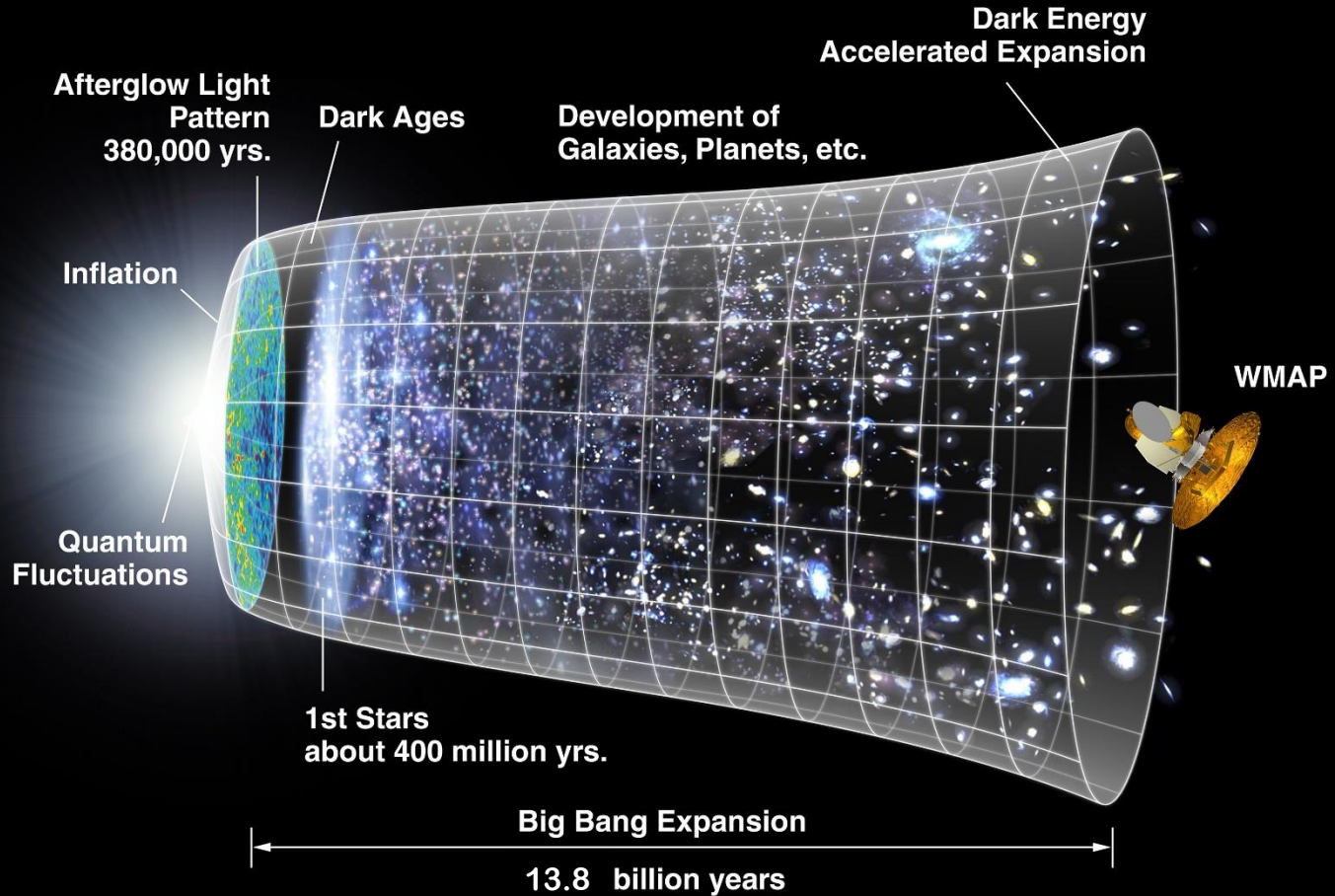
天文学におけるデータ共有 ～デジタル宇宙の構築とその活用～

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。

観測的研究の一般的流れ

- 課題設定・計画立案
- 望遠鏡による観測
- データ処理
 - 校正, 選択, 結合, , ,
- データ解析
 - 物理量の導出
 - **考える**
 - 現象の理解
- 論文出版

データ



情報



知見



理解



(学問の)進歩

天文データの爆発的増加

2015年4月21日

「データシェアリングを利用した科学
技術」に関する勉強会 第4回

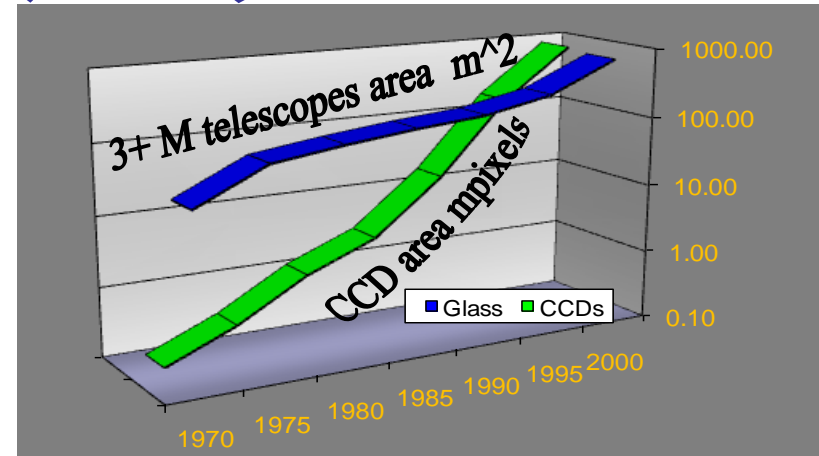
5

急増する天文データ

Courtesy of Alex Szalay

✓ 天文データは毎年倍増

- CPU の計算性能は18カ月で2倍
- I/O 性能の向上率は 10%/年
- 並列計算処理は必須。
- データ移動を極力避ける解析システム



✓ 望遠鏡・観測装置の大型化、高機能化

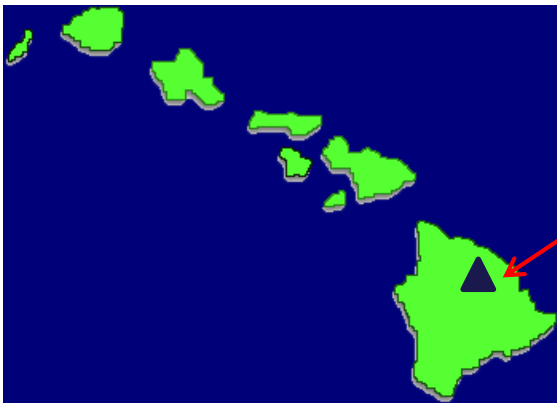
- データ取得のコストの増大 → 科学成果の最大化
- 取得したデータを速やかに解析できる環境を構築し、国内研究者間で共有できる仕組みづくり

✓ 中小望遠鏡との連携観測の進展

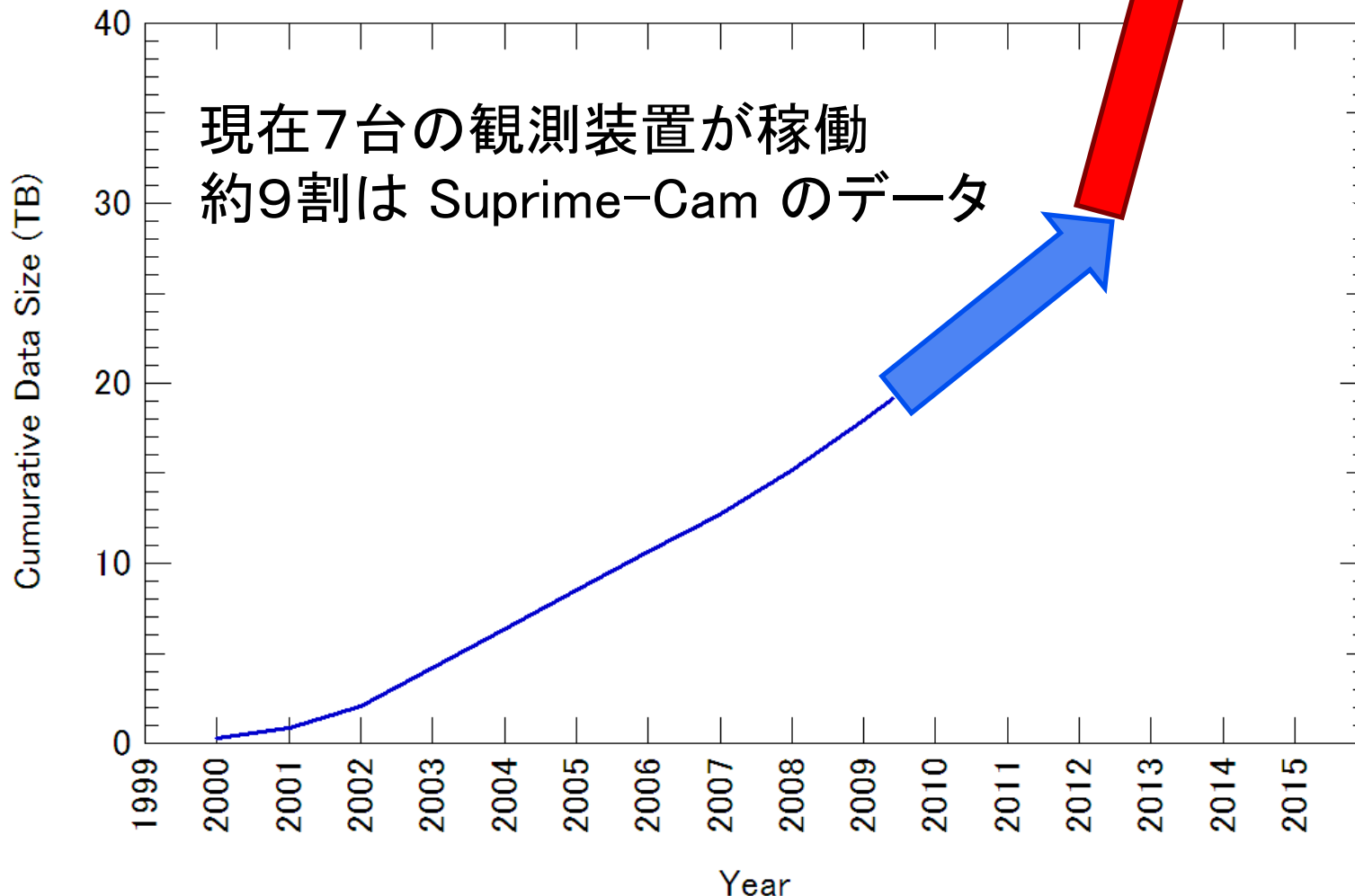
- ガンマー線バースト、AGN の時間変動
- Real time で解析し、結果を速報

すばる望遠鏡

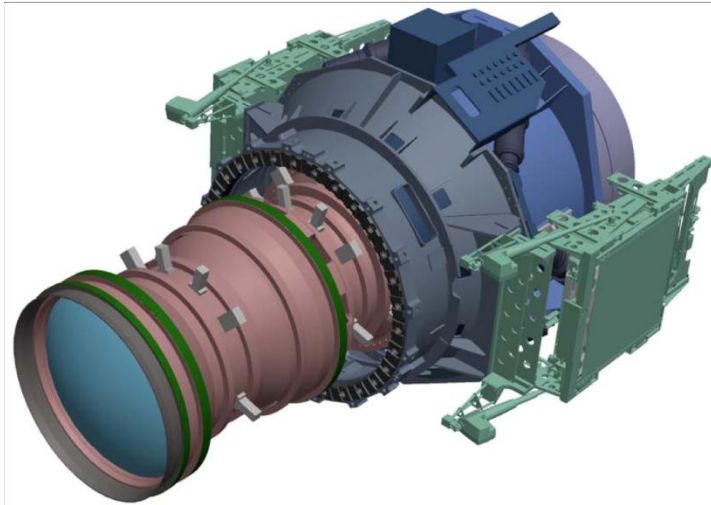
- 世界最大級の光赤外線望遠鏡
- アメリカ・ハワイ島 マウナ・ケア山頂（標高 4,205m）
- 口径 8.2m
- 共同利用装置
- 国立天文台が運用



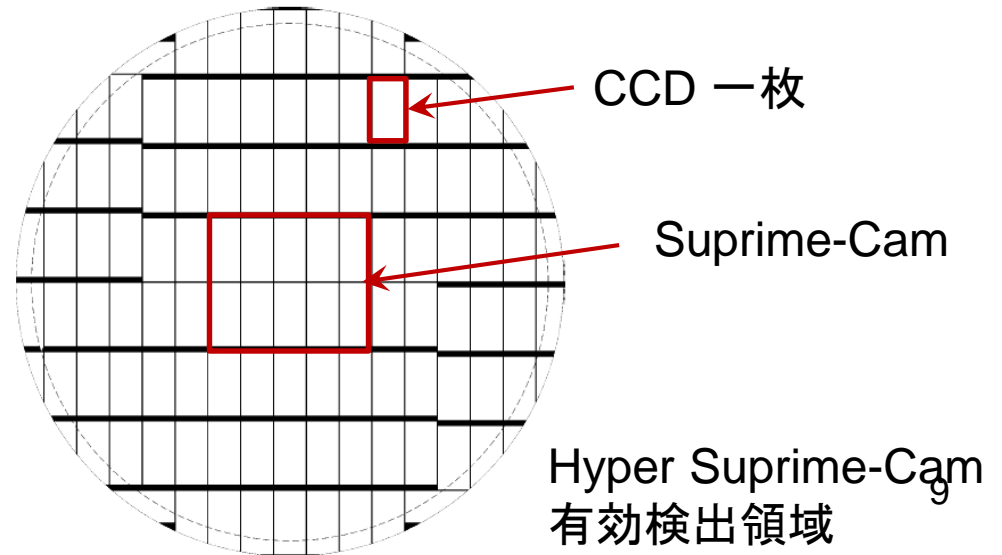
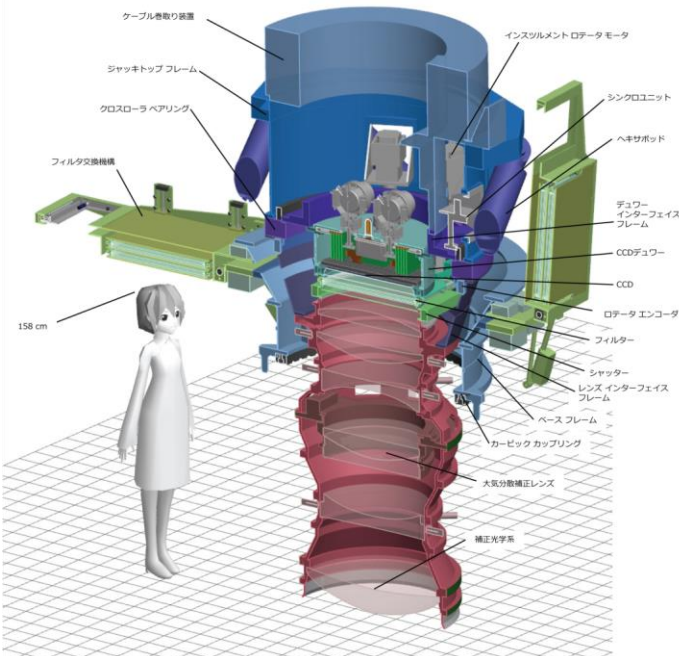
すばる望遠鏡のデータ



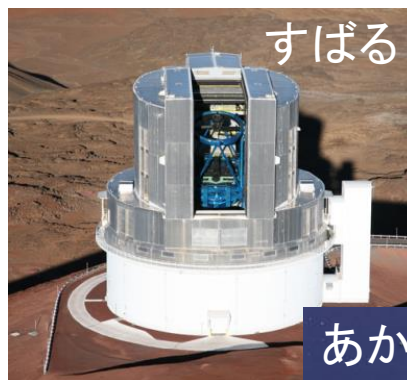
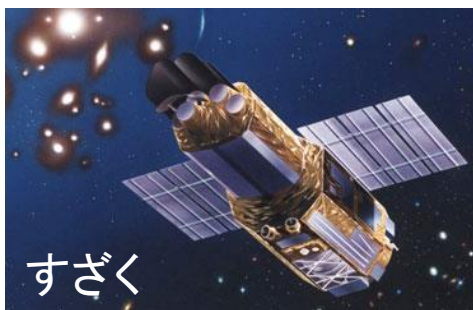
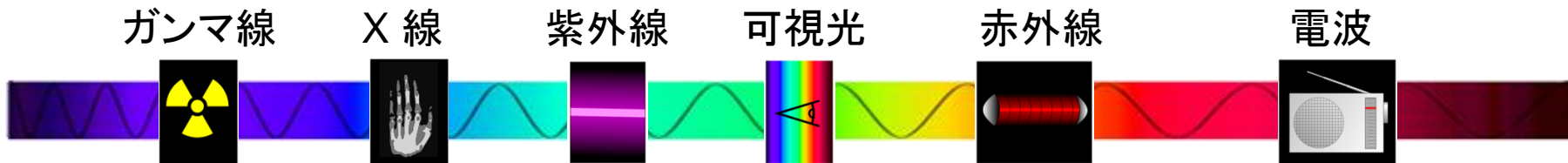
Hyper Suprime-Cam (HSC)



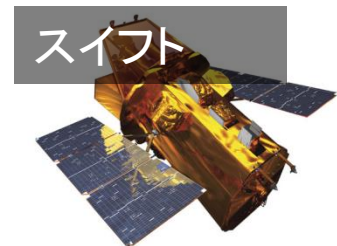
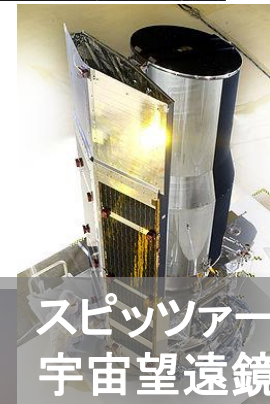
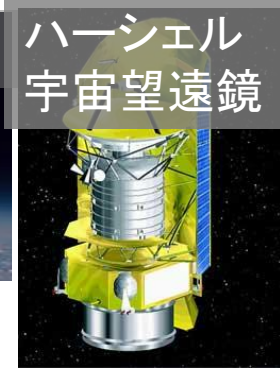
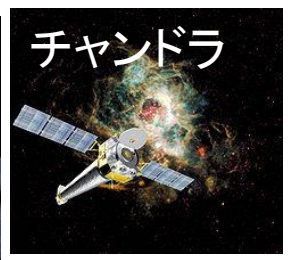
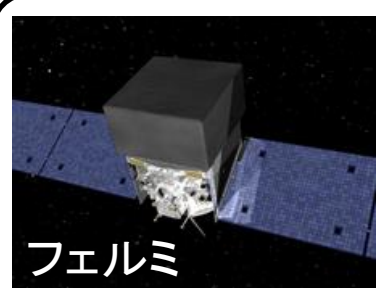
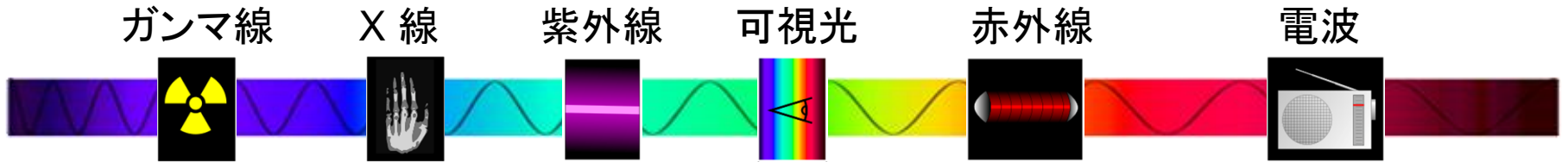
- 2k x 4k CCD を 116 枚搭載
- 800メガピクセルの巨大デジタルカメラ
- 一夜あたり 300~500 GB のデータ生成
- 5年間300夜にわたり、最大 2000 平方度の領域を観測予定 → 計 ~150 TB
- 重力レンズ効果を利用した宇宙のダークマター・ダークエネルギーの解明等
- 2014年春観測開始



稼働中の主要な望遠鏡 (国内)



稼働中の主要な望遠鏡 (海外)

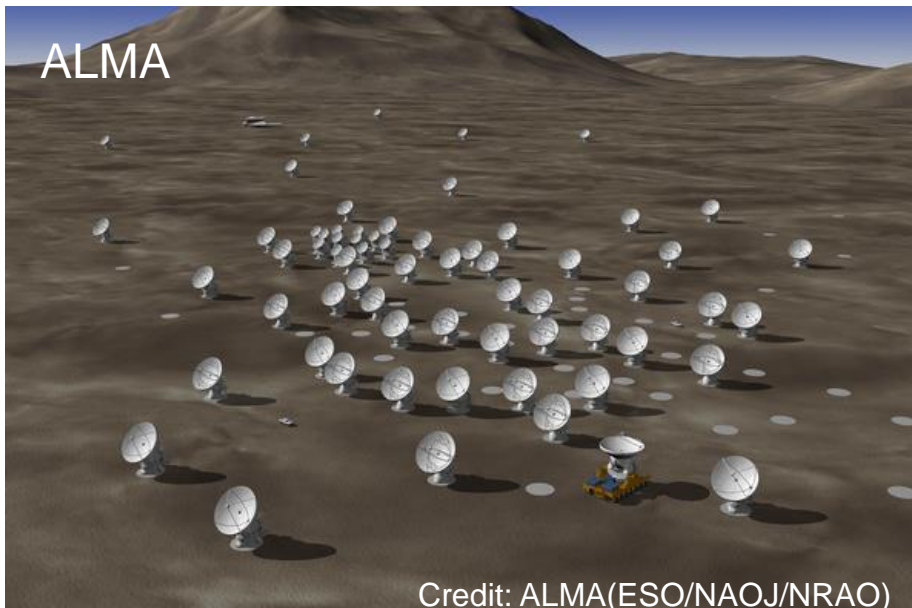


2015年4月21日

「データシェアリングを利用した科学技術」に関する勉強会 第4回

大規模データを生成する望遠鏡計画

ALMA

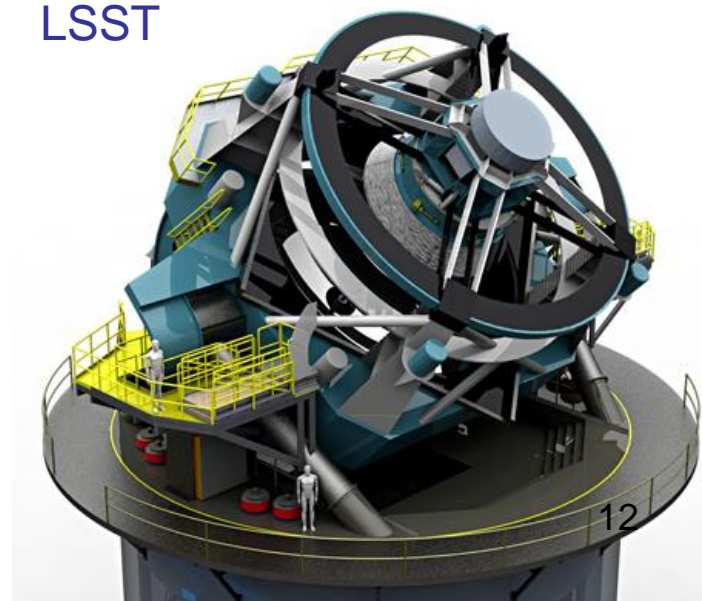


Credit: ALMA(ESO/NAOJ/NRAO)

- 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



天文アーカイブ利用の歴史

明月記

藤原定家の日記
(1180~1235年)

望遠鏡発明前の超新星記録(7件)のうち
3件(1106年、1054年、
1181年)を記録

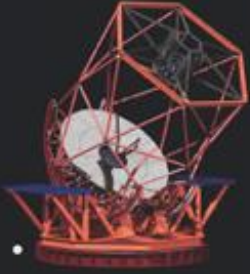
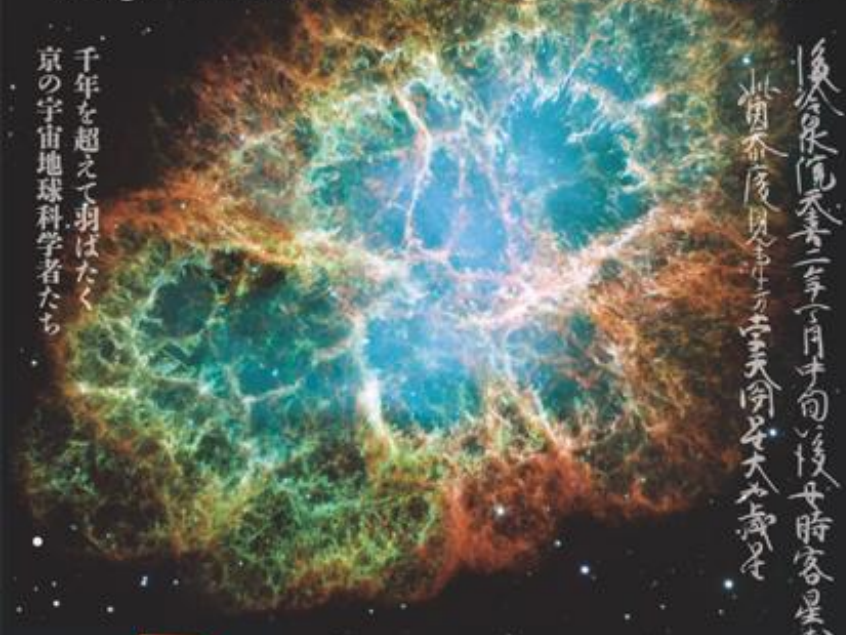
京都大学総合博物館 平成26年度特別展

明月記と最新宇宙像

Meigetsuki and the latest images of our universe

千年を超えて羽ばたく
京の宇宙地球科学者たち

後醍醐天皇 永享三年(1321)月事旬後 母時客星也
此箇亦度見事 万幸天竺大丸歳也



2014年
9月3日[水]—
10月19日[日]

9時30分—16時30分(入館は16時まで)
休館日 ●月曜日・火曜日・曜日・振日にかかわらず
観覧料 ●一般400円 高校生・大学生300円
小学生・中学生200円
※観覧料は16時以降は別途お支払いください。

主催 京都大学総合博物館 京都大学理学部理学系
科学史文芸員 京都大学大学院文学部文学系科学史学専攻
学芸員 京都大学大学院文学部文学系科学史学専攻
学芸員 NPO法人 山崎史学センター 京都大学
宇部総合学芸員 京都大学研究開発センター
協賛 株式会社 山崎史学センター
協力 京都府立総合資料館 白眉科学研究所 京都府立
天文台 山崎史学センター
後援 京都府教育委員会 京都府教育委員会
フーズグループ 協賛

4次元デジタル宇宙ツアー
3D映像で宇宙の奥深い世界を体験してください。
※観覧料は別途お支払いください。

明月記本展期間 9月17日(金)~9月29日(日)
※観覧料は別途お支払いください。

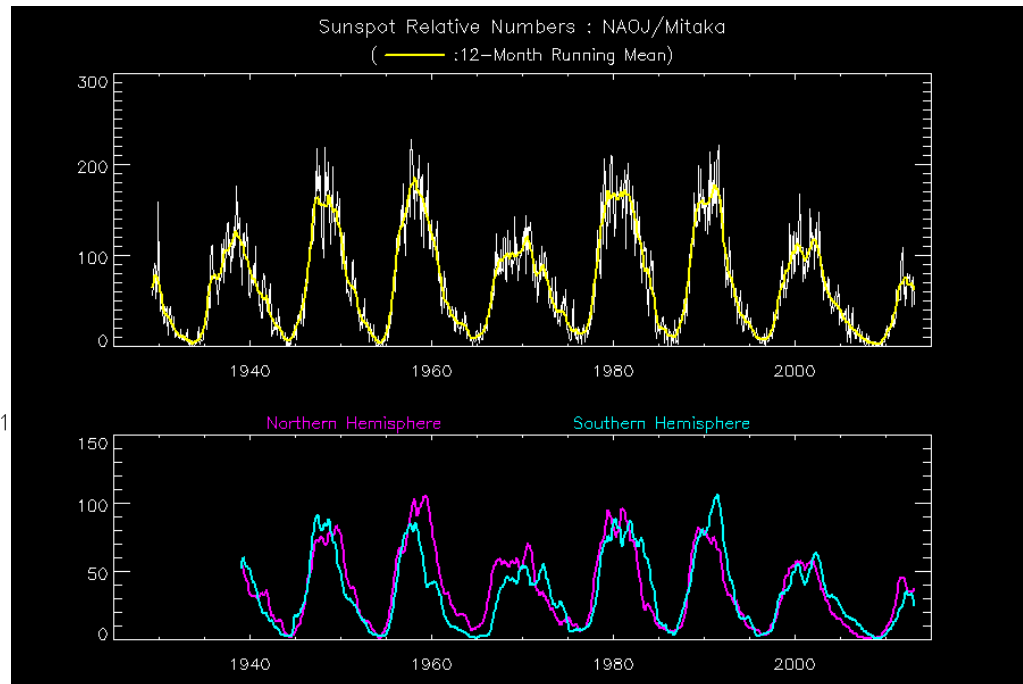
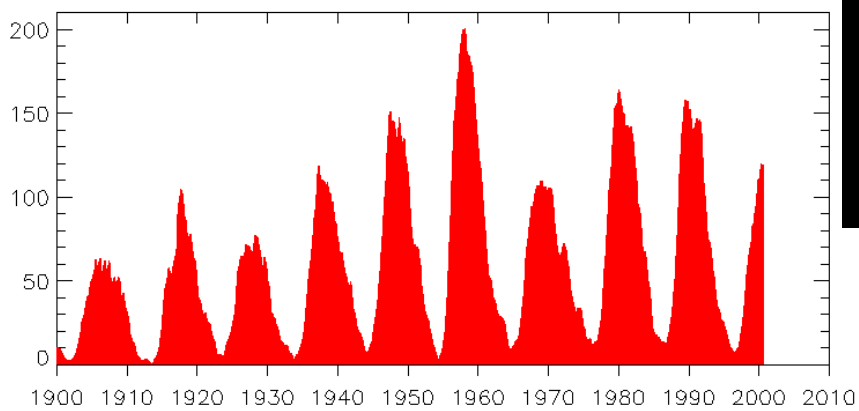
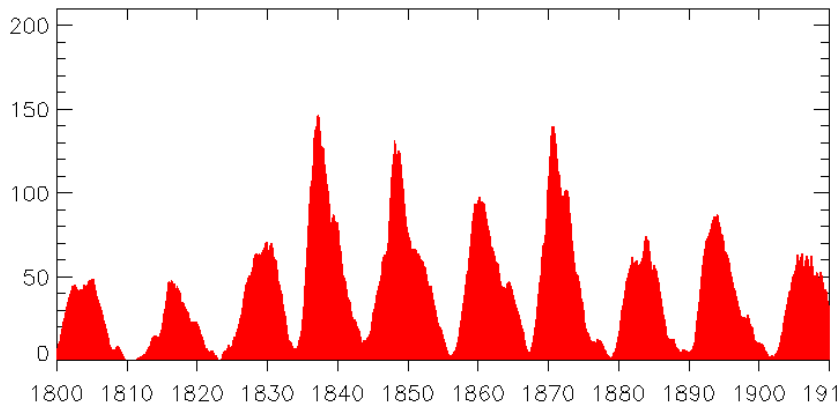
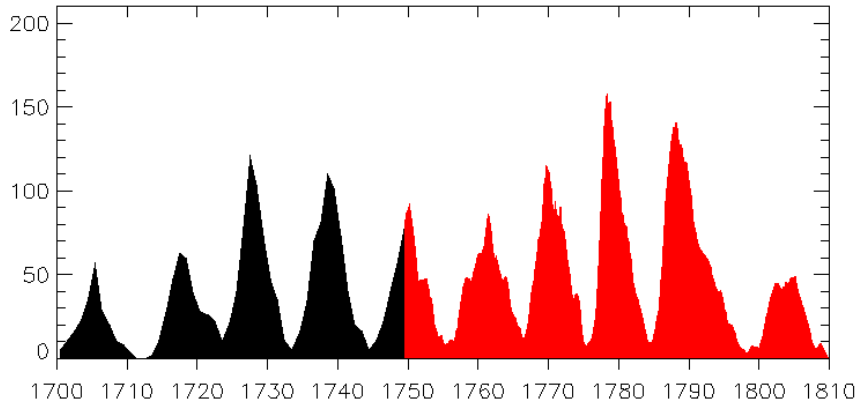
京都大学総合博物館
〒606-8501 京都府京都市古田本町
TEL 075-753-3272
http://www.museum.kyoto-u.ac.jp/

2015年4月21日

「データシェアリングを利用した科学
技術」に関する勉強会 第4回

14

黑点数

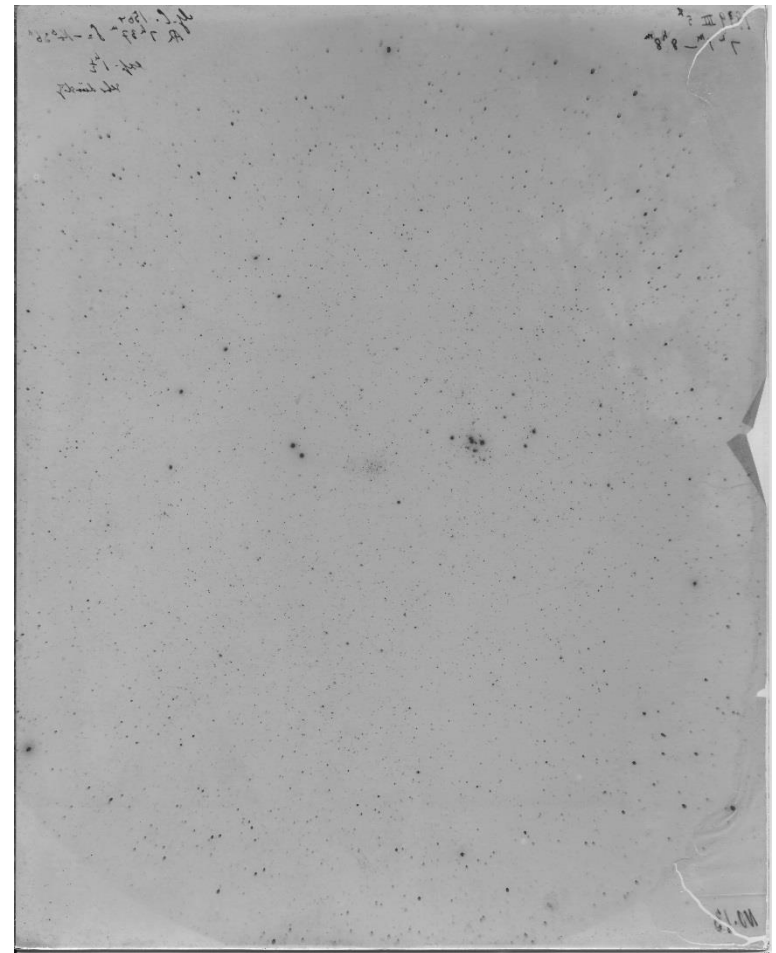


(<http://sidc.oma.be>, Apr 1, 2001)

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)。

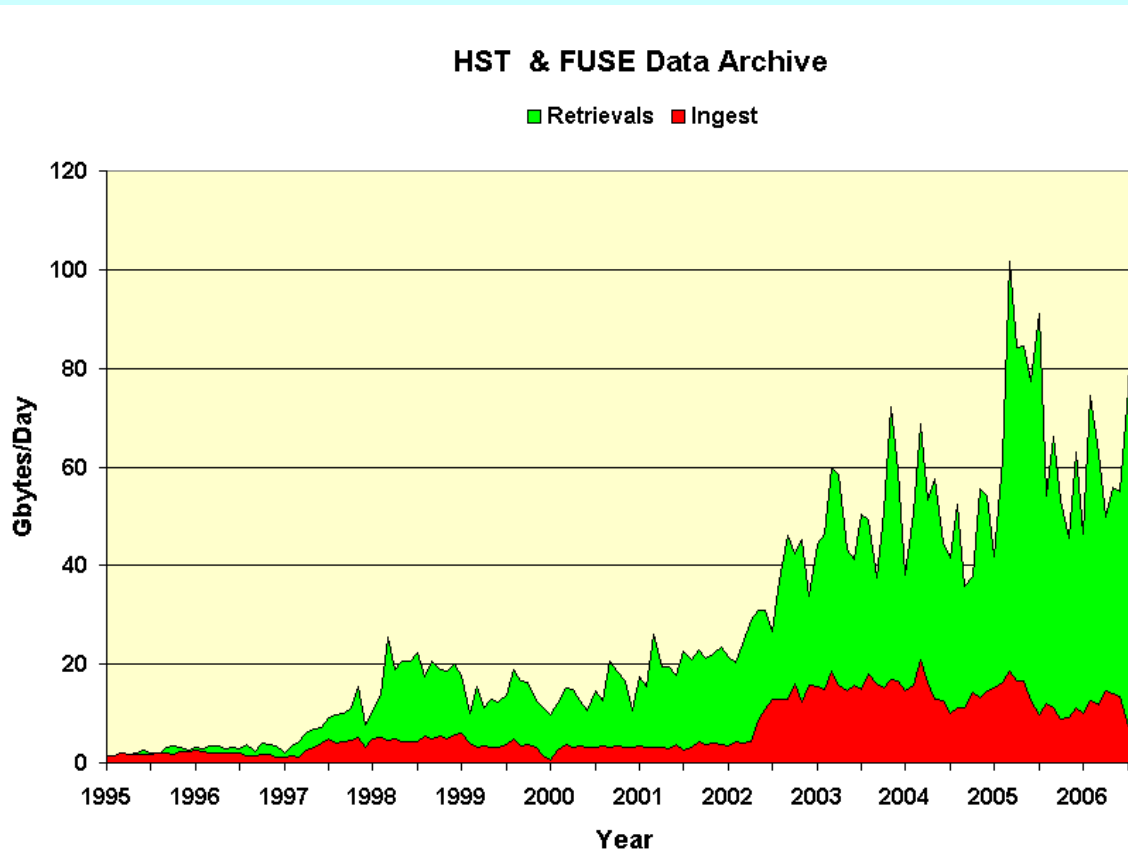


天文学アーカイブの必要性(2)

- 特に最近では、ダイナミックに変化する事象も検出できるようになり (time-domain astronomy)、宇宙の記述にはアーカイビングは不可欠という事情もある。

アーカイブを利用した天文研究

- large fraction of astro-papers based on archives
- HST archive use growing faster than archive



2015年4月21日

技術に関する勉強会 第4回

graphics from
US NVO¹⁸
project

世界のonline天文衛星DBの例

- NASA National Space Science Data Center – COBE, IUEデータの公開
- NASA Goddard Space Flight Center – 高エネルギー天文衛星データの公開
- Infrared Processing and Analysis Center – IRASデータの公開
- Space Telescope Science Institute – HSTデータの公開

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>

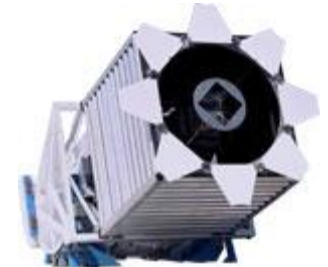
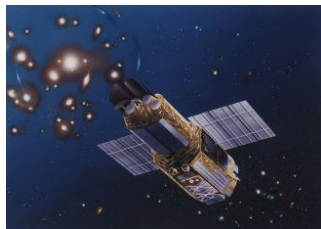
<http://www1.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cadc/>

<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 Astronomique 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

国立天文台が提供する オンラインリソース

[http://www.adc.nao.ac.jp/
Portal/naoj-portal.html](http://www.adc.nao.ac.jp/Portal/naoj-portal.html)

Astronomical Observation
Data Service/ Analysis Portal For Astronomers and Scientists

English

ホーム このサイトの概要 国立天文台 職員向け FAQ (一読必須) その他

提供しているサービス/機能 プロジェクト名をクリックすると概要が確認できます。

光赤外

すばる 観測 データ提供

岡山観測所 観測

SMOKA (天文データセンター) データ提供

太陽

太陽観測所 データ提供 観測

多波長データ解析システム (天文データセンター) 解析

JVO (天文データセンター) データ提供 解析

太陽データ解析システム (天文データセンター) データ提供 解析

ひので科学プロジェクト データ提供 観測

その他

水沢 VERA 観測、他

ALMA 観測 データ提供

野辺山宇宙電波観測所 観測

水沢 RISE 観測、他

電波

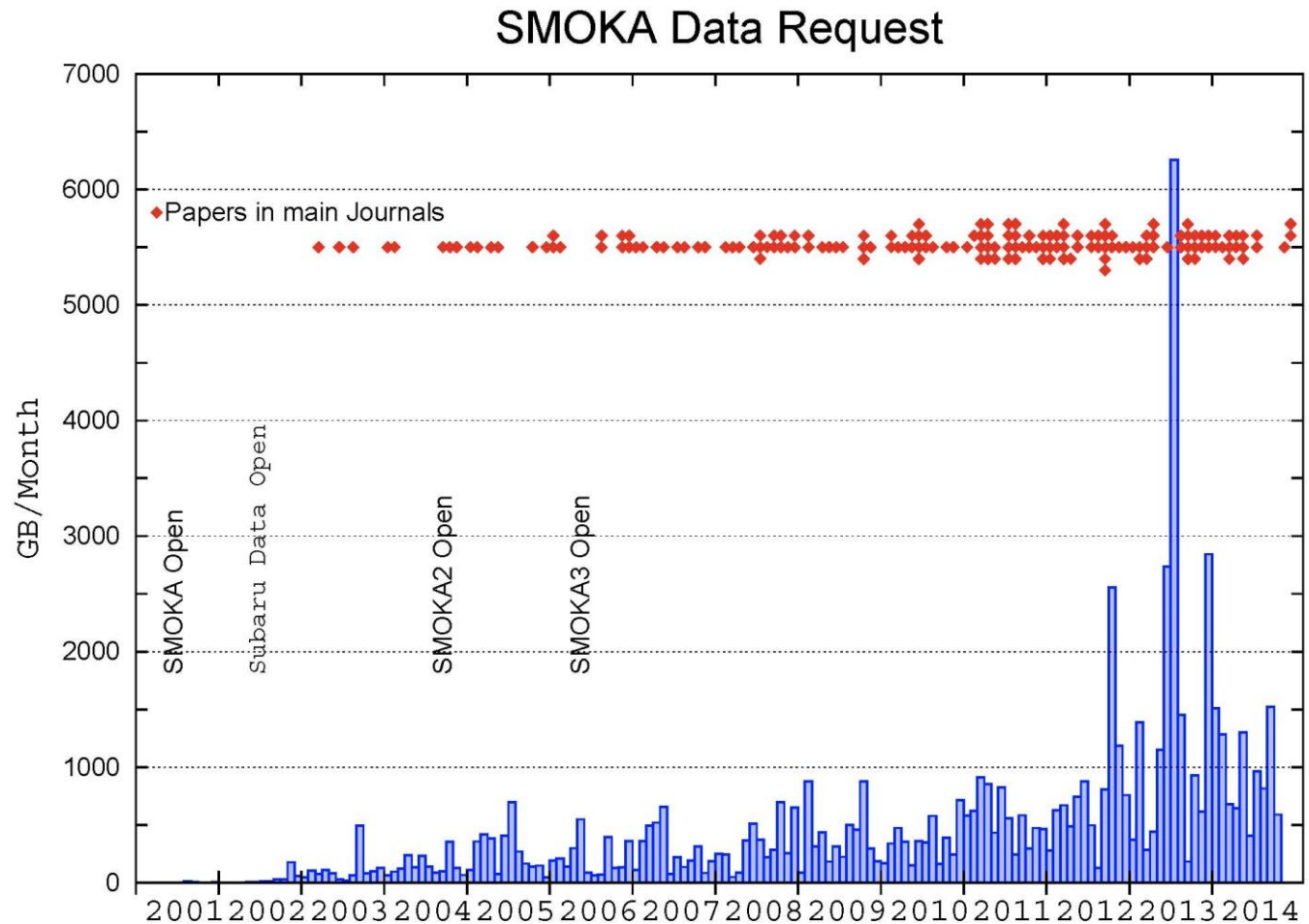
Operated by
Astronomy Data Center (ADC) / National Astronomical Observatory of Japan (NAOJ)

2015年4月21日

「データシェアリングを利用した科学
技術」に関する勉強会 第4回

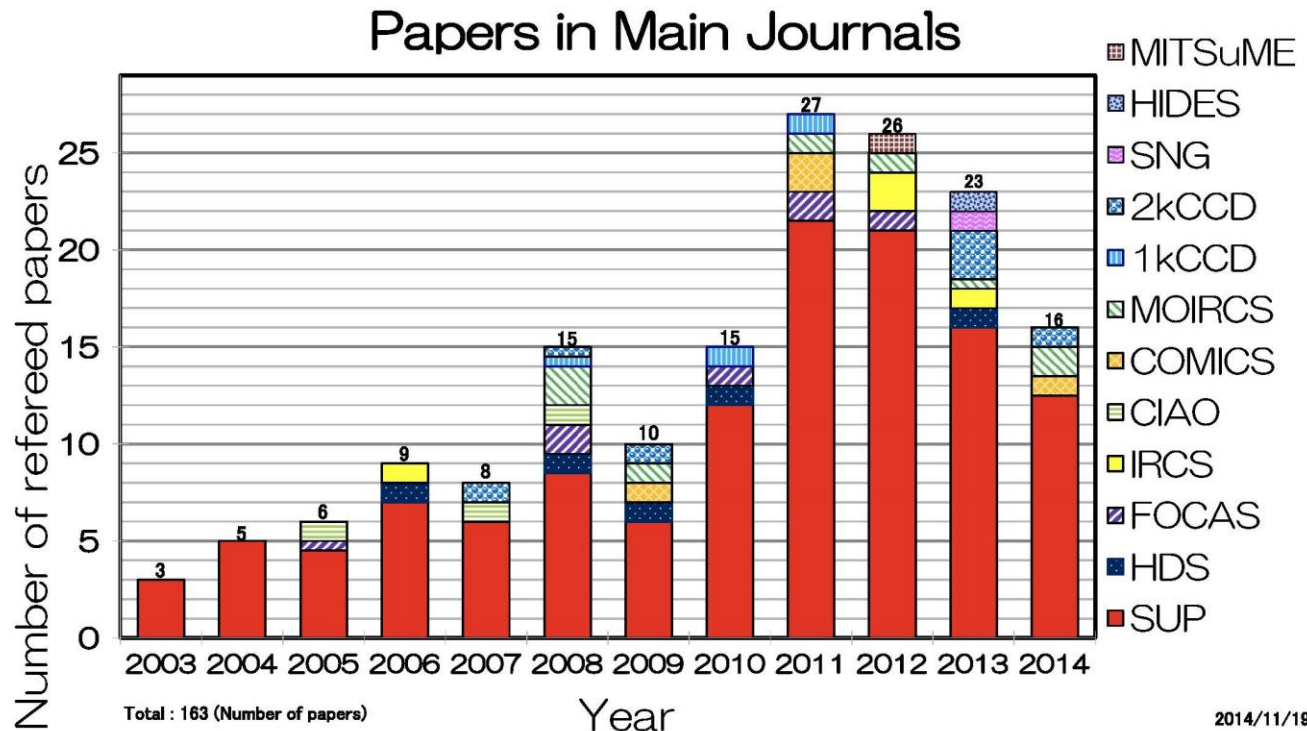
22

SMOKA へのアクセス統計



SMOKAを用いた査読論文

163 papers

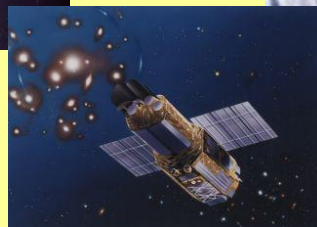
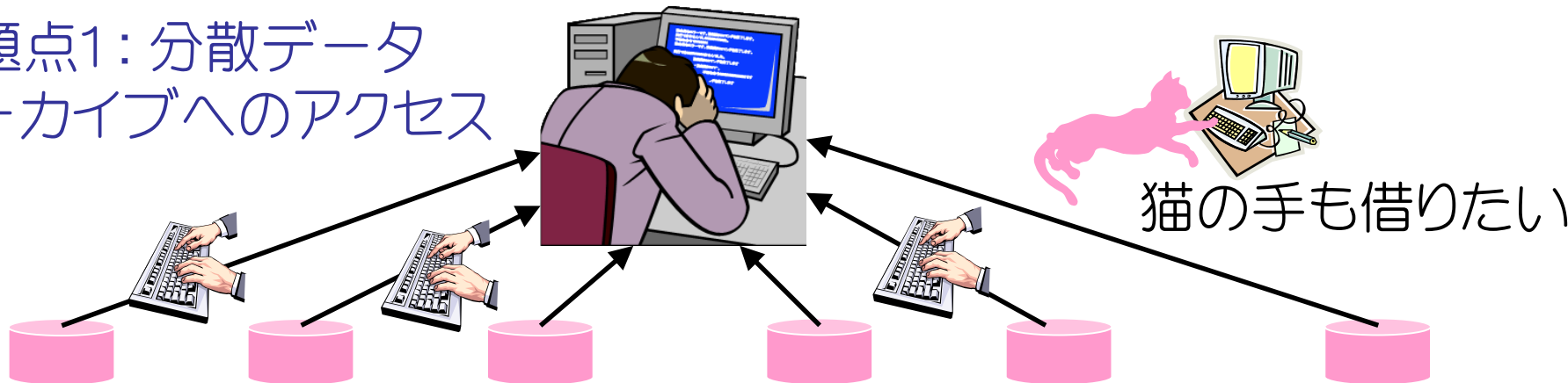


公開・非公開の判断

- 天文分野では、得られた観測データは、提案者による専有期間終了後、原則として公開
- 専有期間の例：
 - すばる：観測終了後18ヶ月
 - アルマ：観測終了後12ヶ月
- 公開目的
 - 観測データの二次利用を促進を通じた**科学的成果の最大化**
 - 高い観測時間獲得競争(例：アルマは最大10倍)に対応

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

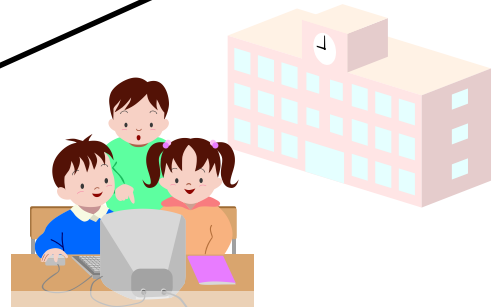


データベースアクセスインターフェースの共通化

データ
取得を
自動化

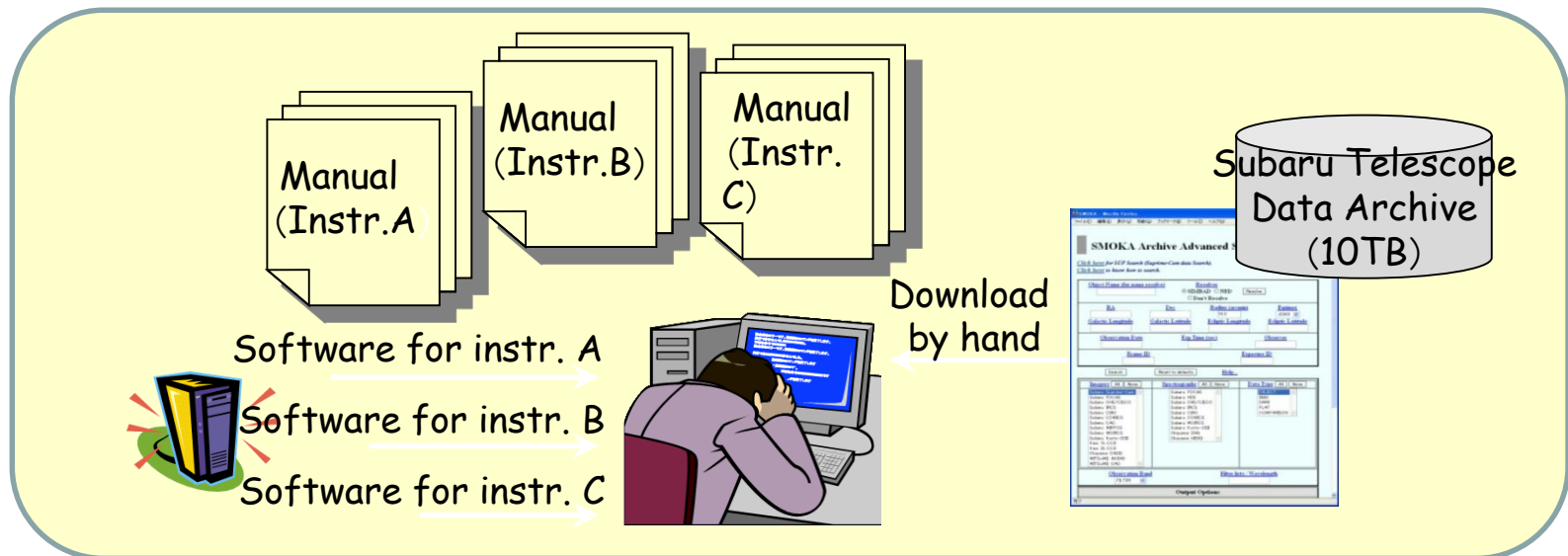


教育用教材としても利用可⁷



問題点2：巨大観測データの処理

- ・ データアーカイブからのダウンロード困難
- ・ 大容量ディスク＋並列計算システムが必要
- ・ 生データのリダクションは装置毎に異なりラーニングコストが高い



ヴァーチャル天文台

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

VO Table 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

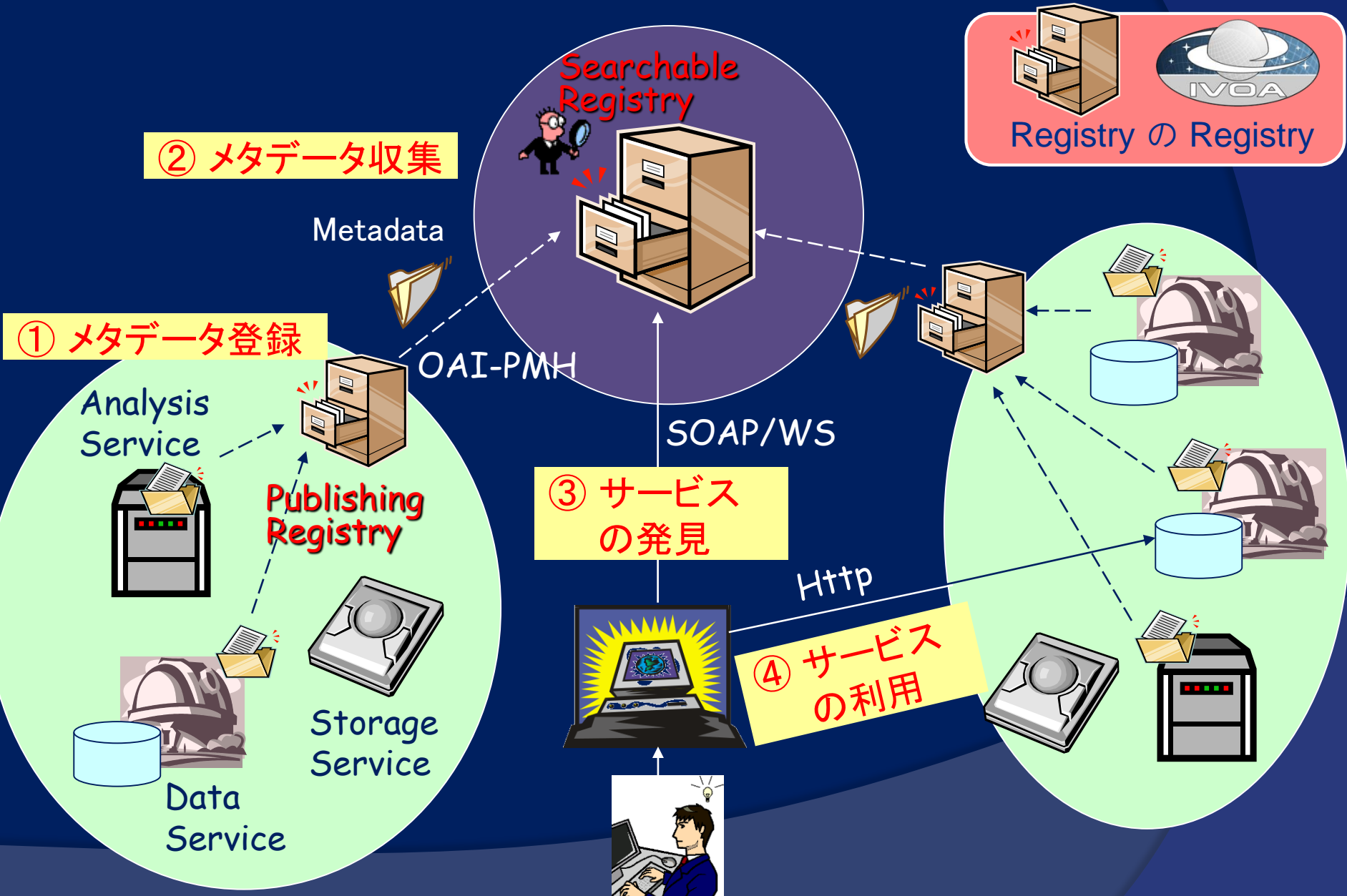
IVOA Astronomical Data Query Language – 検索言語仕様

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

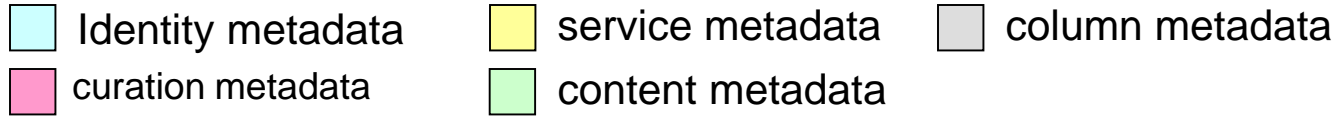
Data Model for Astronomical DataSet Characterisation – データモデル

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



Resource Metadata



title	string
short_name	string
identifier	URI
publisher	string
publisher_id	URI
creator	string
creator_logo	URL
contributer	string
date	string
version	string
contact_name	string
contact_email	e-mail address
service_interface_url	URL
service_base_url	URL
service_http_result	MIME type
service_standard_uri	URI
service_standard_url	URL
service_msr	float,decimal degrees
ucd	string
unit	string
datatype	string
width	int
precision	string
arraysize	string

subject	string,list
description	string,free text
source	string
reference_url	URL
type	string,list
coverage_spatial	string
coverage_region_of_regard	float,decimal degrees
coverage_spectral	string,list
coverage_spectral_bandpass	string,list
coverage_spectral_central_wavelength	float
coverage_spectral_minimum_wavelength	float
coverage_spectral_maximum_wavelength	float
coverage_temporal_start_time	string
coverage_temporal_stop_time	string
coverage_depth	float
coverage_depth_unit	string
coverage_object_density	float
coverage_object_count	int
coverage_sky_fraction	float
resolution_spatial	float
resolution_spectral	float
resolution_temporal	float
content_level	string,list
facility	string,list
instrument	string,list
format	string,list
right	string

	identity	curation	service	content	column
catalog	○	○	○	○	×
table	○	○	○	○	×
column	○	○	○	○	○

データ形式

- Flexible Image Transfer System (FITS)
 - 観測データを磁気テープに書き込み、持ち運ぶための標準。1981年に初版が策定された。
 - 1 record = 2880 bytes
 - (Header, Data)(Header, Data)...
 - 国際天文学連合に FITS WG があり、仕様を保守。
- VOTable
 - ヴァーチャル天文台での標準データ形式
 - XMLを用いた半構造データで IVOAで標準化
 - FITSファイルを内部に包含可能、または、FITS ファイルへのURLを保持。

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)

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

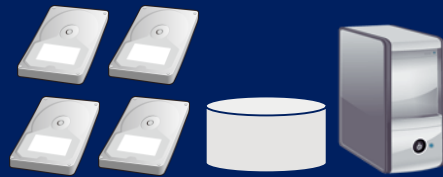
JVO system の概略

五つのサブシステムと
分散データサービス (VO Service)

JVO Computing GRID

CUI for jc (jvo command)

JVO Portal



Subaru Archive

JVO Registry

Registry



VO Service



JVOSpace

An example for using the
computing service

- ① Job submit
- ② Submit to JVO grid
- ③ Save result on JVOSpace
- ④ Retrieve the result

An example for using the VO
query service

- ① Submit a query
- ② Search a VO service
- ③ Search to VO service
- ④ Retrieve a FITS image

②

①

①

②

③

④

④

③

JVO ポータル (GUI) の機能

データ検索

1. 高速検索

JVO 内部のデータベース
主要なカタログを登録

2. サービスを一つ指定して検索

サービスを検索し選択
テーブル選択
検索条件指定

3. 複数サービスへの同時検索

領域指定検索のみサポート

4. JVO Sky サービス

Google Sky IF を利用したデータ検索

データサービス検索

1. キーワード検索

2. カテゴリ検索

3. 詳細検索

すばるデータ検索・解析

1. Suprime-Cam

モザイク画像作成

フラットフレーム作成

2. HDS

代表的なサービス・ツール

✓ Web ポータル

- JVOポータル <http://jvo.nao.ac.jp/portal>
- NVOポータル <http://www.us-vo.org/>

✓ アプリケーション

- Aladin : 天体画像表示
- TOPCAT : グラフ作成
- VOPlot : グラフ作成
- VOSpec : スペクトルデータ表示・解析

✓ その他

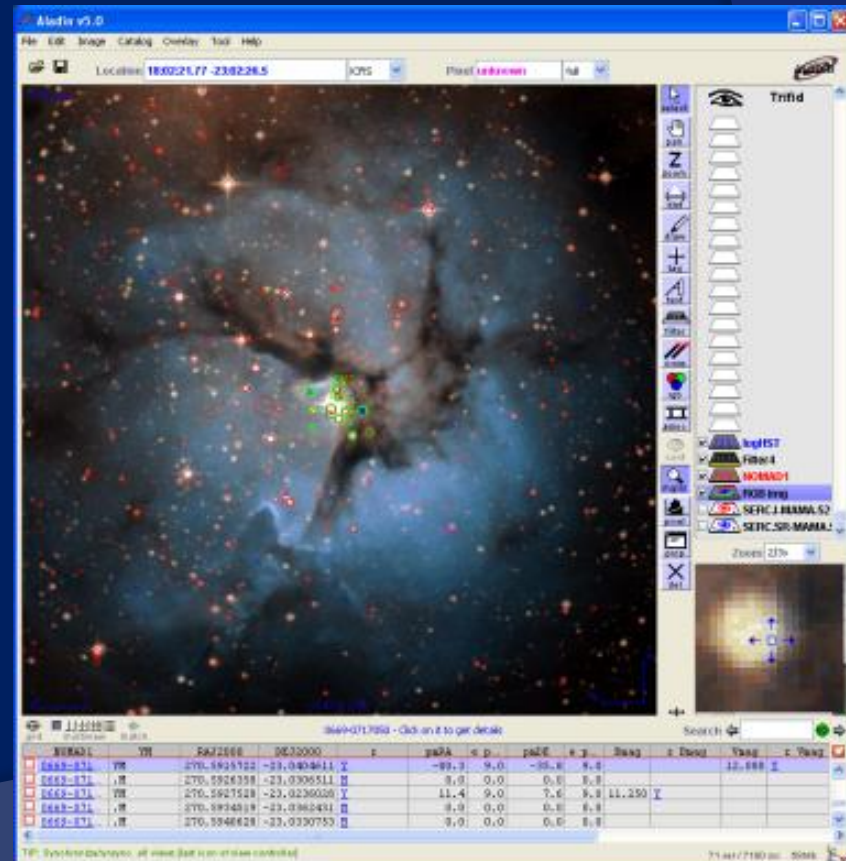
- <http://www.ivoa.net/cgi-bin/twiki/bin/view/IVOA/IvoaApplications>
- <http://www.euro-vo.org/pub/fc/software.html>

Aladin



<http://aladin.u-strasbg.fr/aladin.gml>

- ✓ フランス ストラスブルグ天文データセンター (CDS)
- ✓ 画像データビューア、天体カタログデータも表示
- ✓ 三色合成
- ✓ 複数カタログ クロスマッチ
- ✓ VOクライアント機能
 - VOサービスから画像・カタログデータを取得して表示
 - さまざまな波長の画像データやカタログデータを横断的に取得可能
 - VOアプリケーションの中で最も使用頻度が高い。



Topcat



<http://www.starlink.ac.uk/topcat/>

✓ イギリス Starlink プロジェクト

✓ 様々な種類のプロット作成

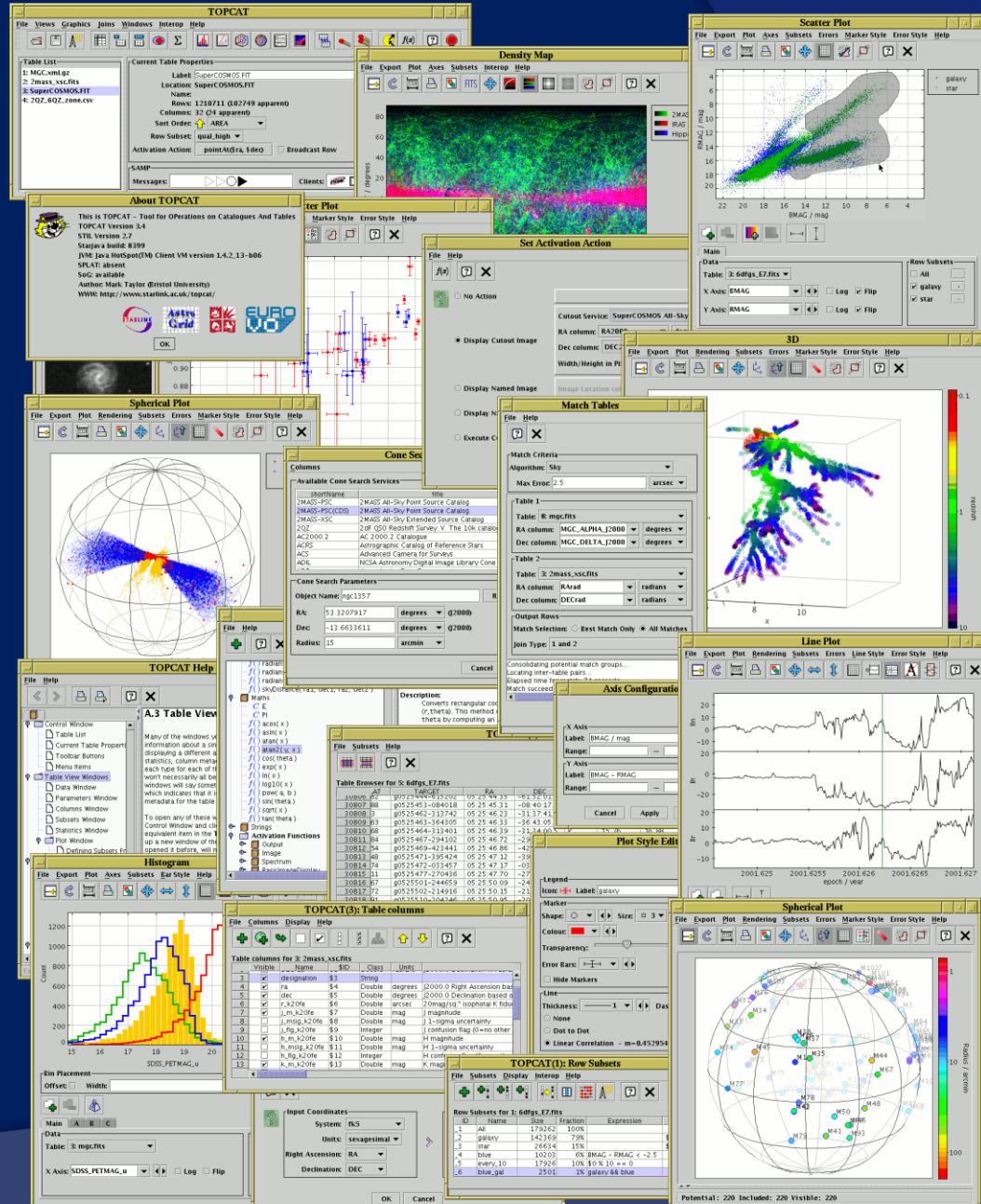
✓ 複数カタログの相互比較機能

✓ VOクライアント機能

— VO カタログデータサービスからデータを取得し、表示。

— Aladin に次ぐ。

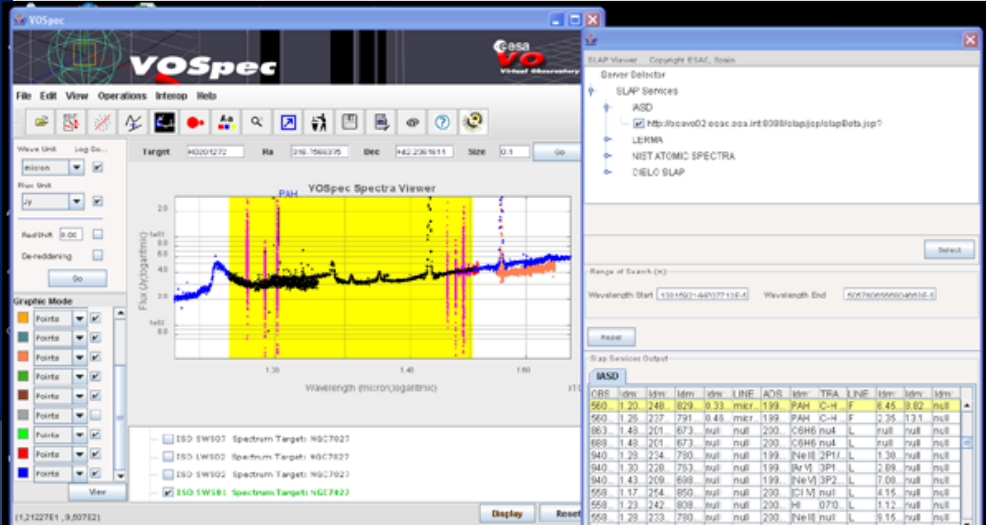
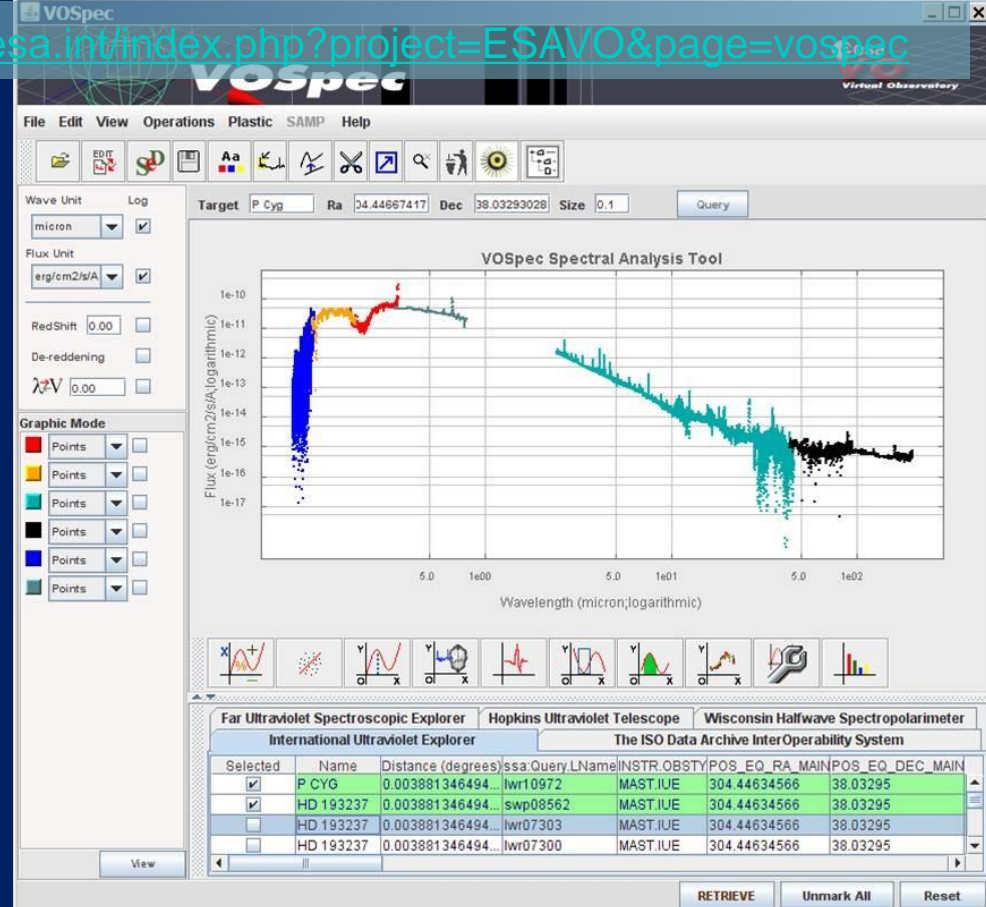
— Aladin とデータ連携。



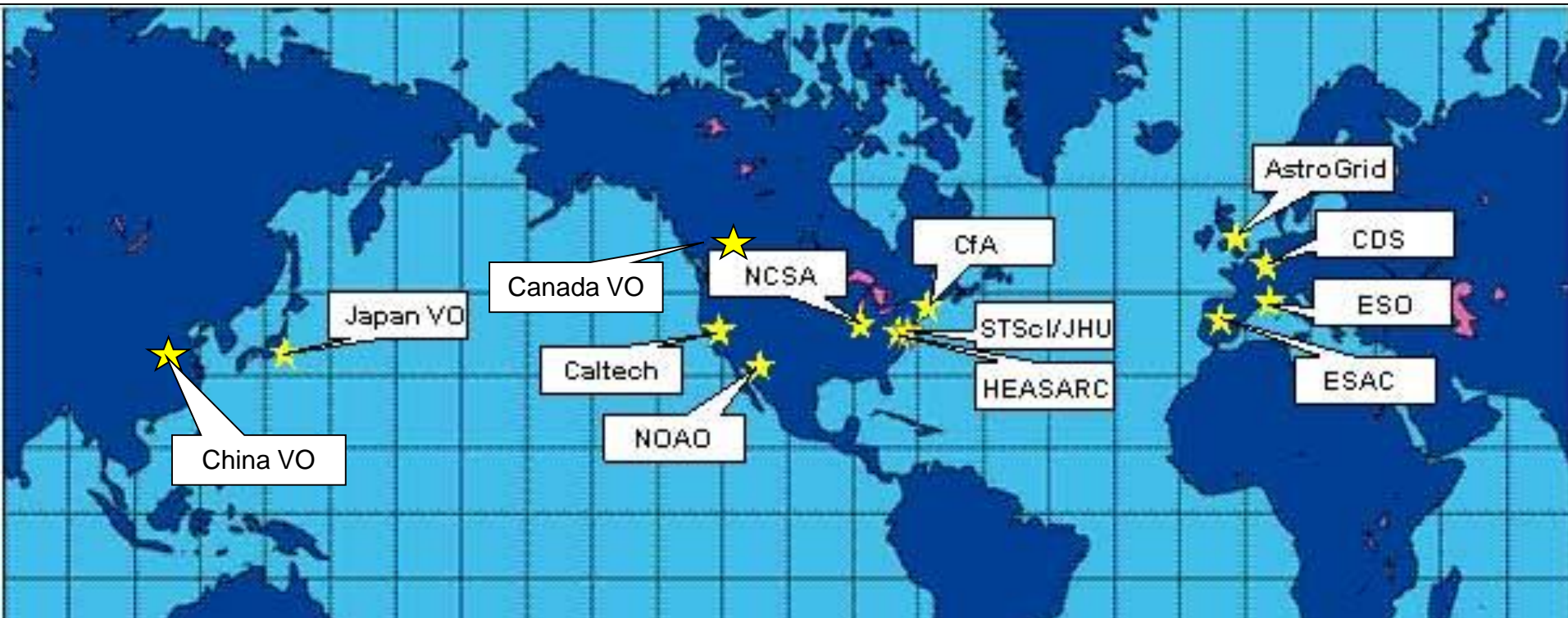
VOSpec

<http://www.sciops.esa.int/index.php?project=ESAVO&page=vospec>

- ✓ ESA/ESAC の VO グループが開発
- ✓ VO データサービスからスペクトルデータを取得できる。
- ✓ 複数のスペクトルデータを重ねて表示
- ✓ 単位変換を行う機能をもつ。単位の異なるスペクトルデータも同時に表示可能
- ✓ フィッティング機能
- ✓ スペクトル線データベースと連携し、ライン同定する機能あり。



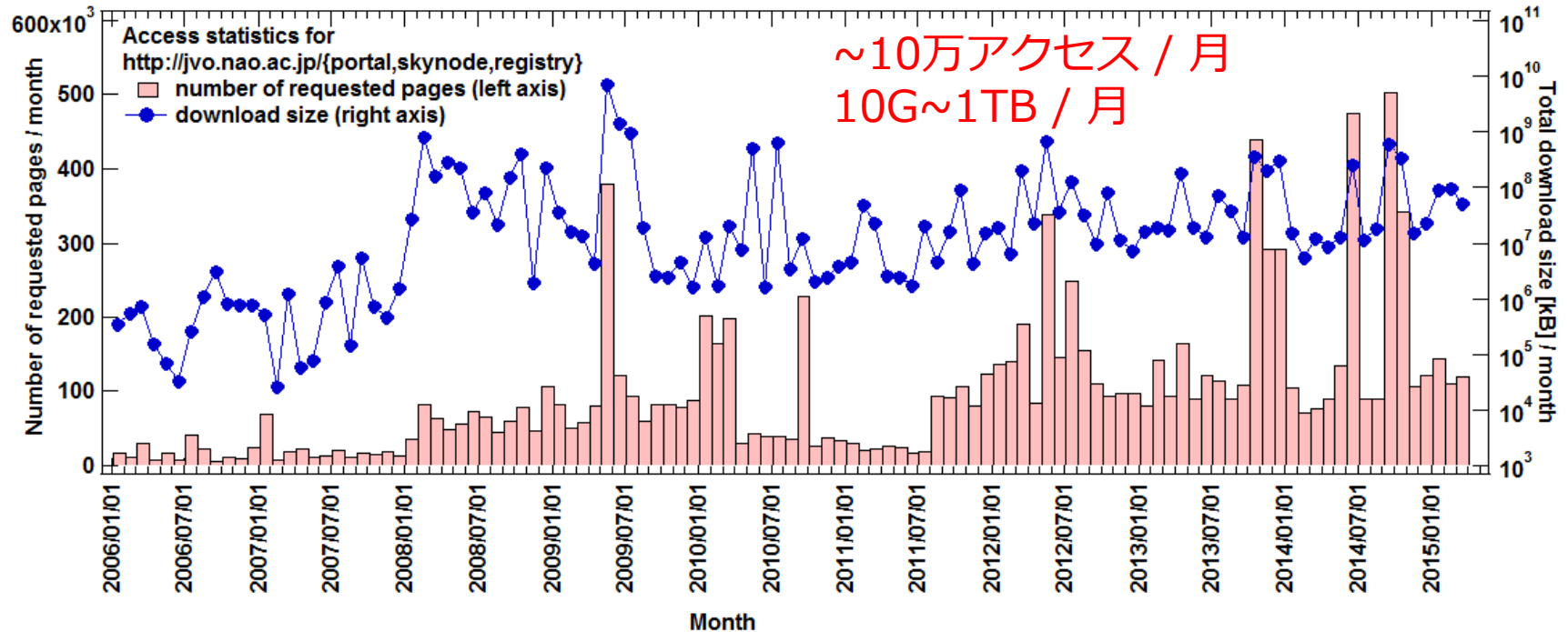
VOにより連携する世界の天文台



More than 16,000 resources are accessible.

Images, spectra, and catalog data can be retrieved

JVOシステムアクセス統計



2014年度

235万アクセス

1.55 TB ダウンロード

Portal 81万アクセス
0.28TB

VO-IF 148万アクセス
1.27TB

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

VOを使って生み出された
査読論文
1883 (ADS調べ)

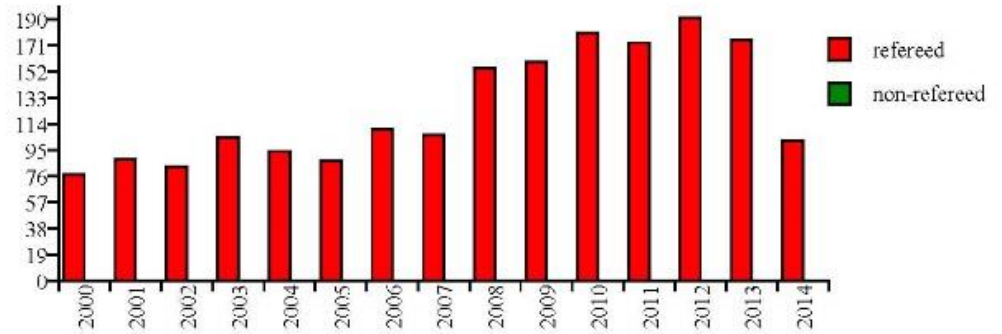
2014年11月現在

READ10 index

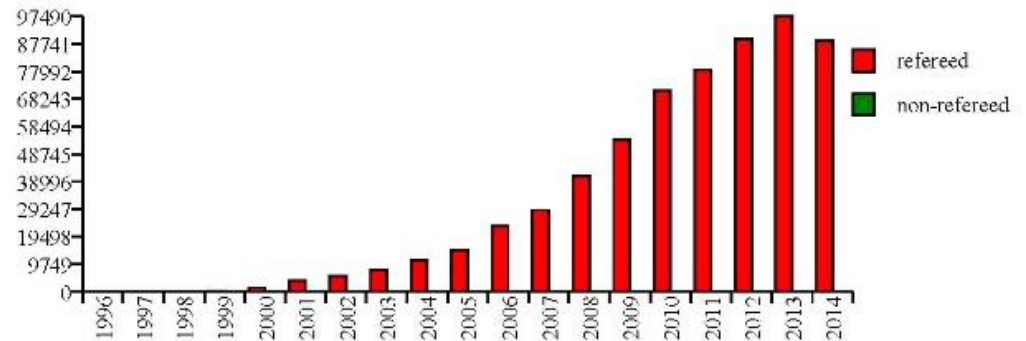
19845

19845

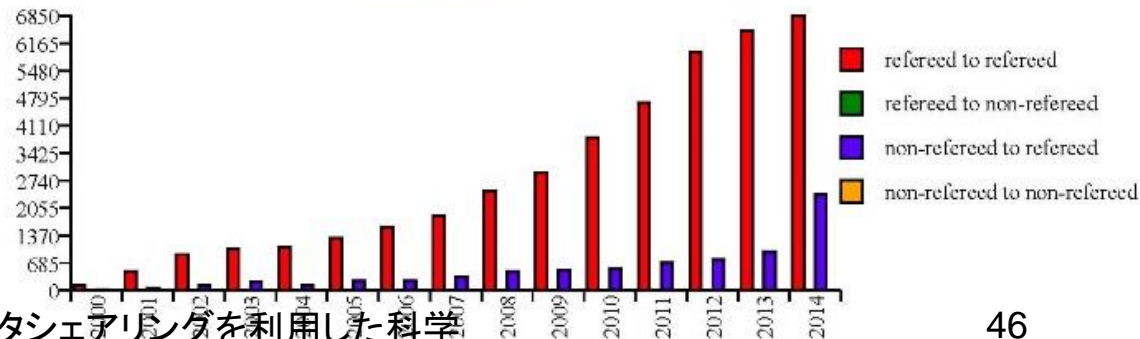
publication histogram



reads histogram



citation histogram



2015年4月21日

「データシェアリングを利用した科学技術」に関する勉強会 第4回

将来に向けて

IAU GA Beijing August, 2012

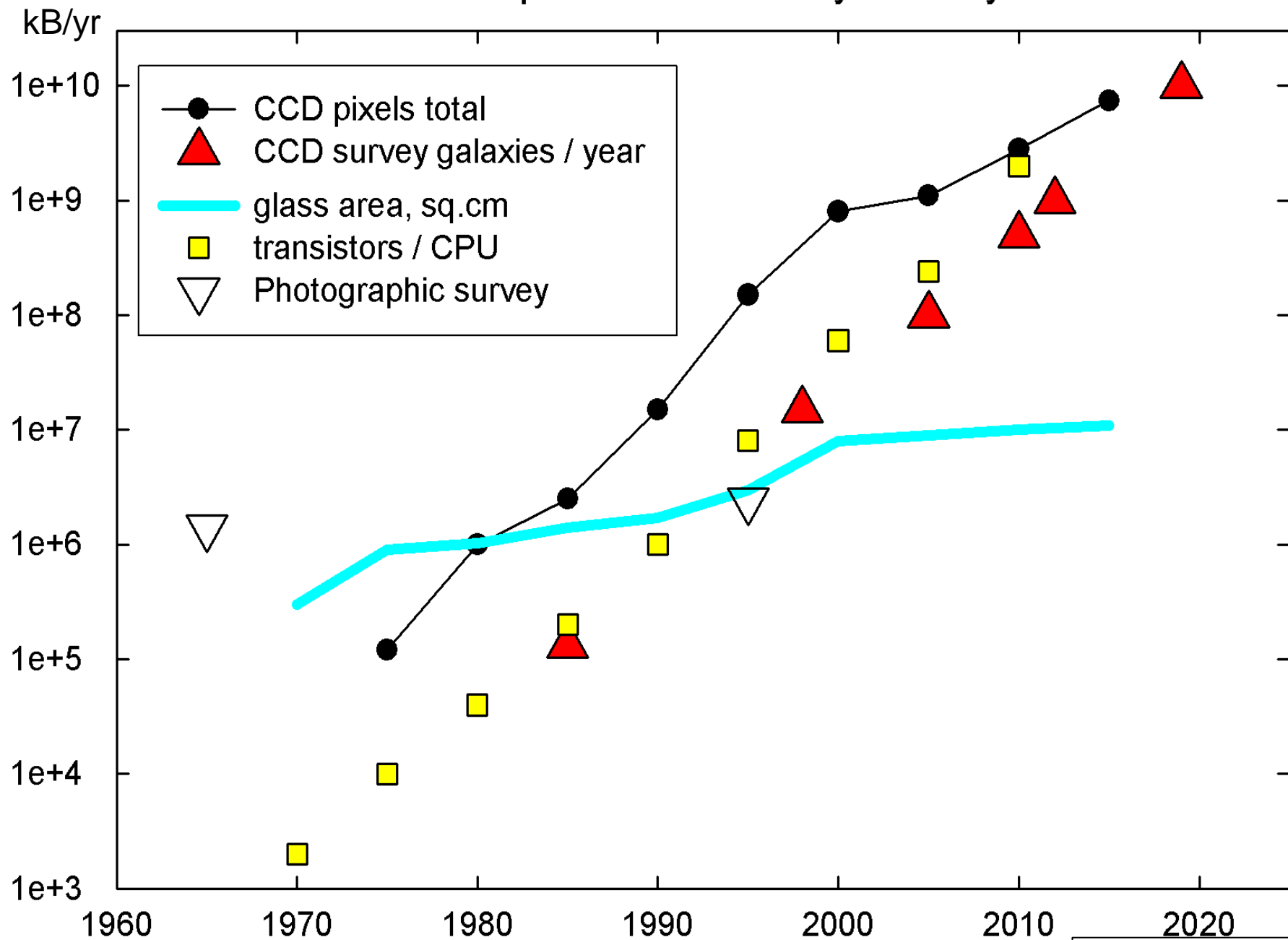
Special Session 15

Data Intensive Astronomy

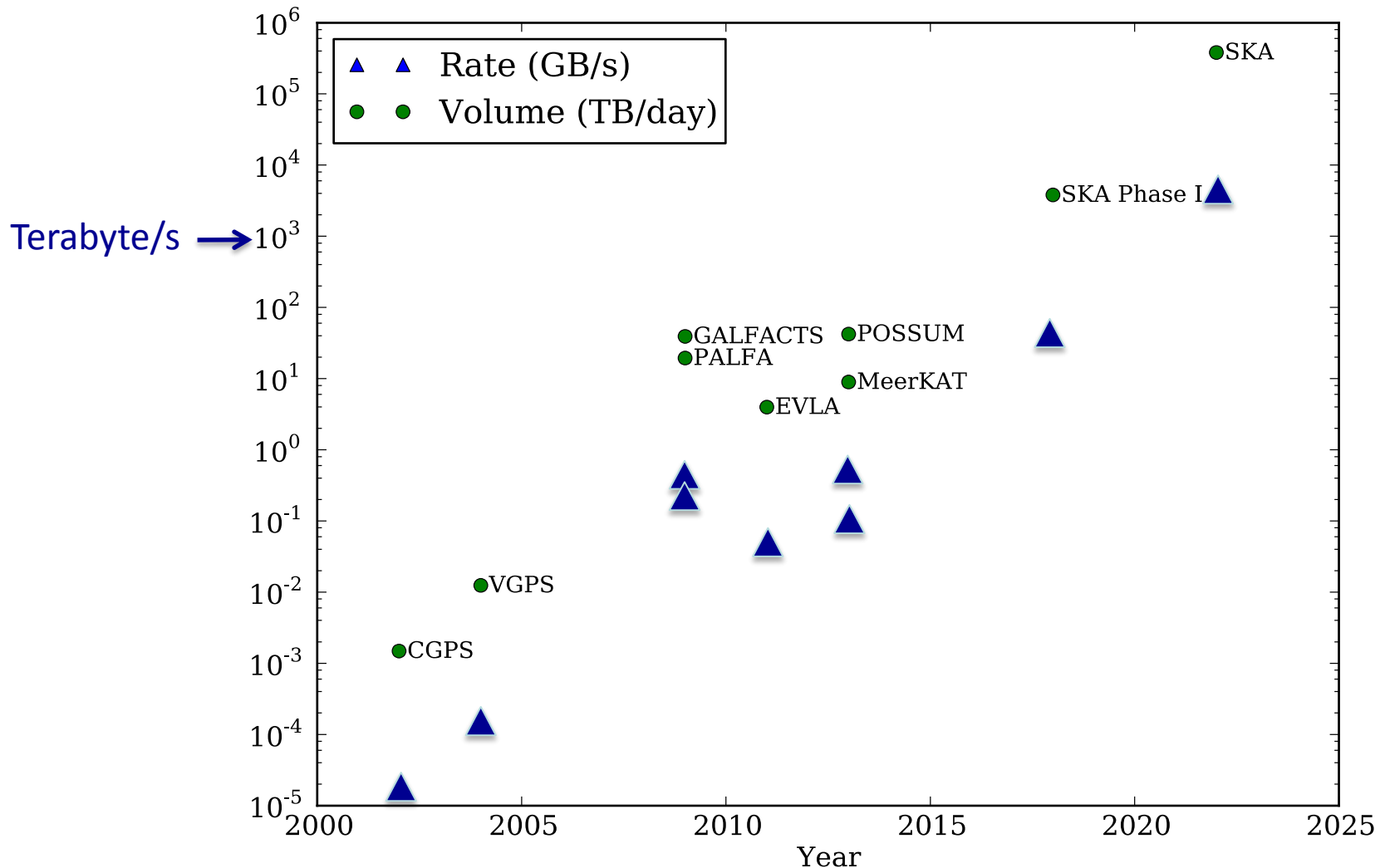
- Scientific Impact of Past and On-Going Large-Scale Observations and Surveys to Astronomy
- Current Status and Challenges of Future Large-Scale Observations and Surveys (1) Near- and Mid-future projects, (2) Far-future projects
- Data Management and Data Access: Past, Present and Future
- Advanced Data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



Trends in Optical Astronomy Survey Data



Survey Raw Data Output Rates



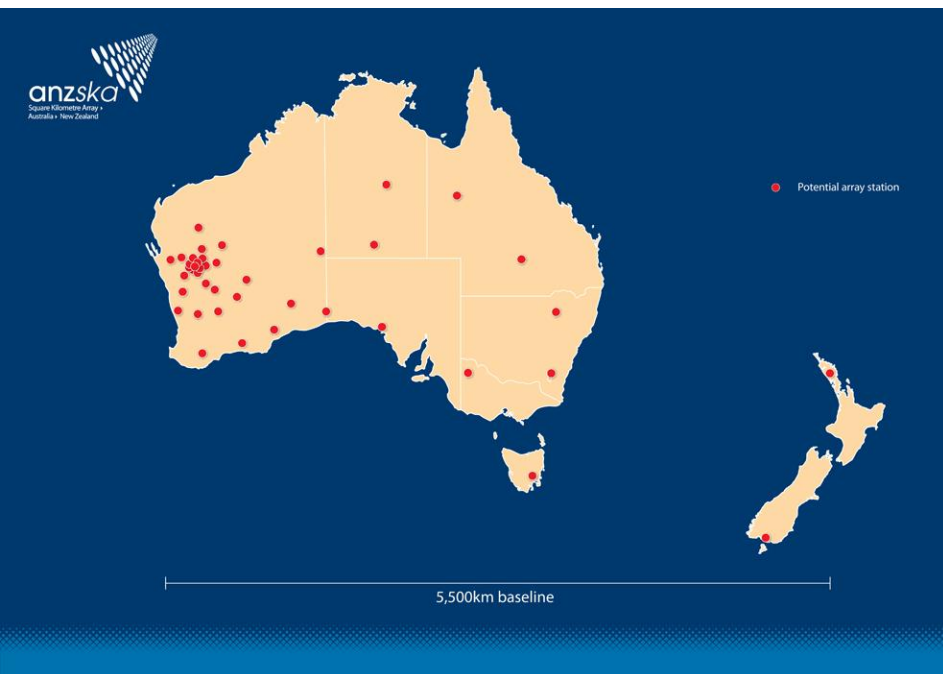
2015年4月21日

「データシェアリングを利用した科学技術」に関する勉強会 第4回

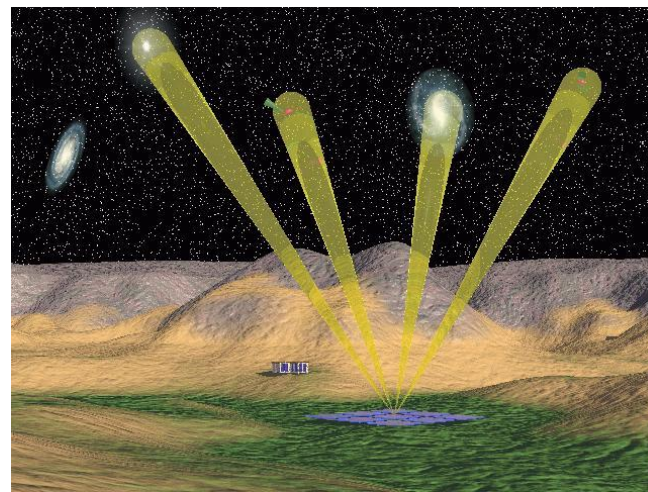
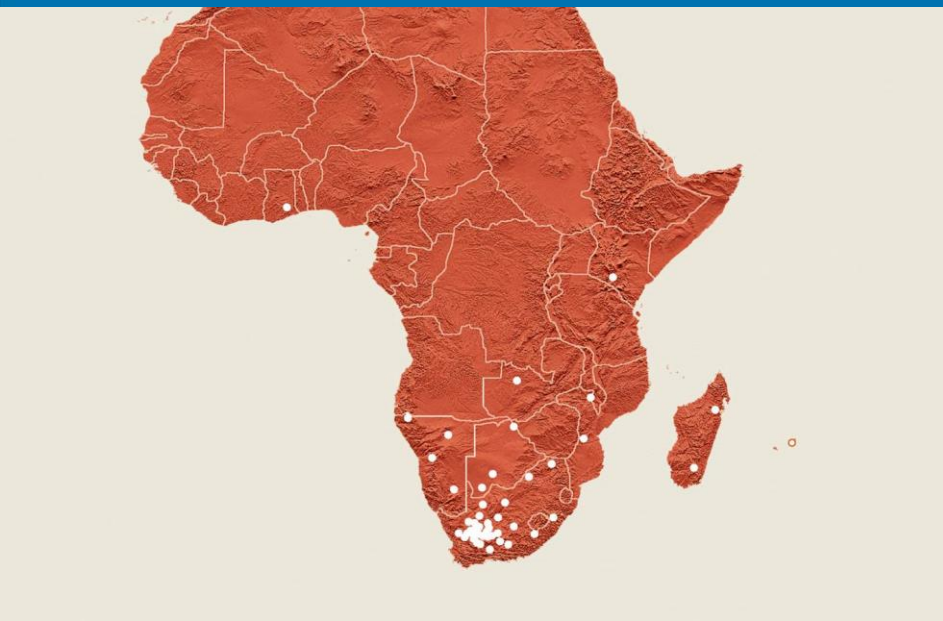
50

Courtesy of R. Taylor

Square Kilometre Array (SKA)



- 集光面積 1km^2
- 開口合成電波干渉計
- 完成~2022 ??
- 豪 & 南ア
 - ASKAP & MeerKAT
 - “1% SKA” プロトタイプ



超大量データ時代の天文学

- 観測データ, シミュレーションデータのオンラインデータベースへのアクセス
 - 多波長天文学, 観測と理論の比較
 - 本当の意味での統計的データ処理
 - 大量データからの新知見の発見
- オンラインデータ解析, データマイニング, etc.

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

Special Challenges in Next Generation Astronomy

- HPCによる高スループットデータ解析
- トップ100スパコンによる大量データストリーミング処理
- データは巨大。しかし、データの寿命は未定義。
 - ALMAのデータならなんとかなりそう
- SKAの場合: スケールする解が必要
ASKAP (1%) → SKA1 (10%) → SKA2 (100%)
- 現行アルゴリズムは順次処理ベースで、数100～数千コアではスケールしない
- 予算は限定される。電力消費はファクター10-100は落とさないと。

データは使われて初めて生きる

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
- **天文が参照例の一つに**

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



2015年4月21日

「データシェアリングを利用した科学技術」に関する勉強会 第4回

56

データシェアリングの成功事例

1. Protein Data Banks (PDB)
2. OneGeology/CGI model
3. Intergovernmental Panel of Climate Change (IPCC)
4. International Virtual Observatory (IVOA)

ICSU Recommendations

- WDSを通じたデータ公開に向けた14事項を勧告
<http://www.icsu.org/publications/reports-and-reviews/strategic-coordinating-committee-on-information-and-data-report>
- オープンアクセス; データポリシーの確立と公表; データマネージメントの重要性; 標準アクセス法やツールの活用; 等
- 持続的データ公開の重要性を指摘

行政に期待すること

- 行政官も研究者も、「**データ公開**」は、ハードウェア製作と並んで、**科学的成果最大化にとって必須**という意識を持つ。
- 永続的なデータ公開を保証するため、分野の中核的研究機関に「**データセンター**」の設置を義務づけ、**予算と人員を措置し、永続するための体制作りを行政が支援**。
- ある程度大型の研究費配分の際に「**公開に耐えるデータ作成**」を強く**推奨**（評価項目にする；c.f., 米国NSFは大型プロジェクトにはデータ公開を義務づけている）
→ 分野の中核的研究機関に譲渡してデータを公開
- データ利用者には、データ作成者や公開機関への**クレジット（謝辞）記載を義務づける**。クレジットはデータ作成者や公開機関の実績にカウントするシステムの確立を。