

大規模データ時代を見据えた データ解析システム

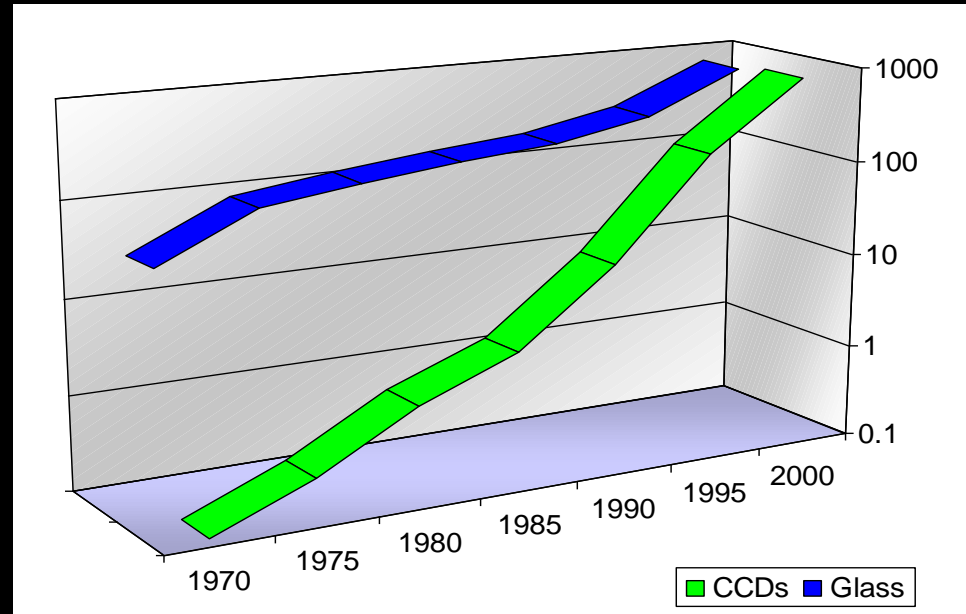
JVO による Suprime-Cam データモザイクシステムの
現状と TMT 時代へ向けた今後の戦略

国立天文台 天文データセンター
白崎裕治

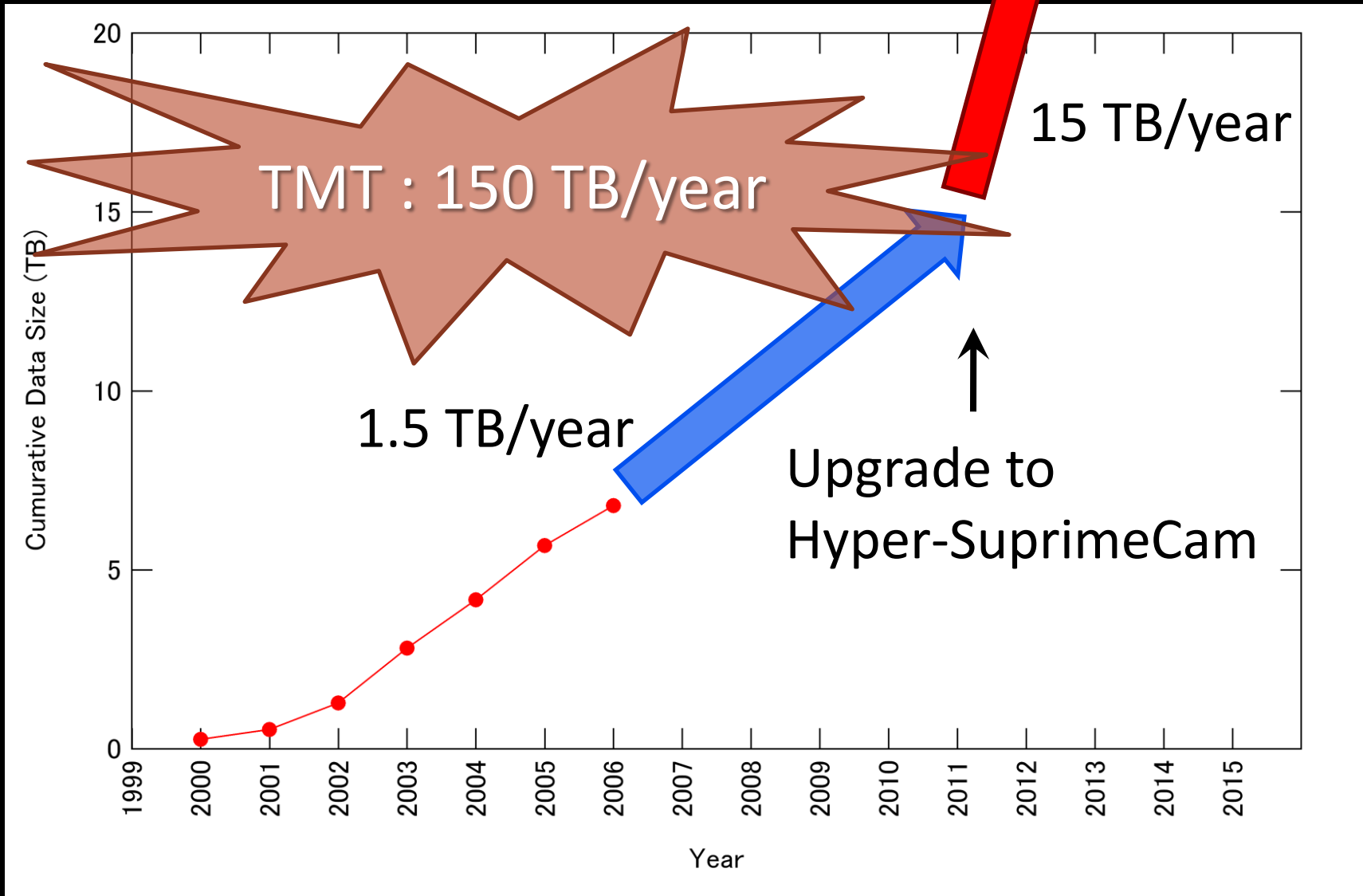
急増する天文データ

- 全世界の天文データは1年で倍増
- CPU 性能が2倍に向上するのに要する期間：2年
- 1 CPU コアあたりの性能向上は限界。マルチコア化。
- 並列実行しないと解析速度は速くならない → 並列計算システム

• ネットワーク速度の向上もデータ量の増大に追いつけていない → データと解析資源を一極集中するのが Best.



すばるのデータ

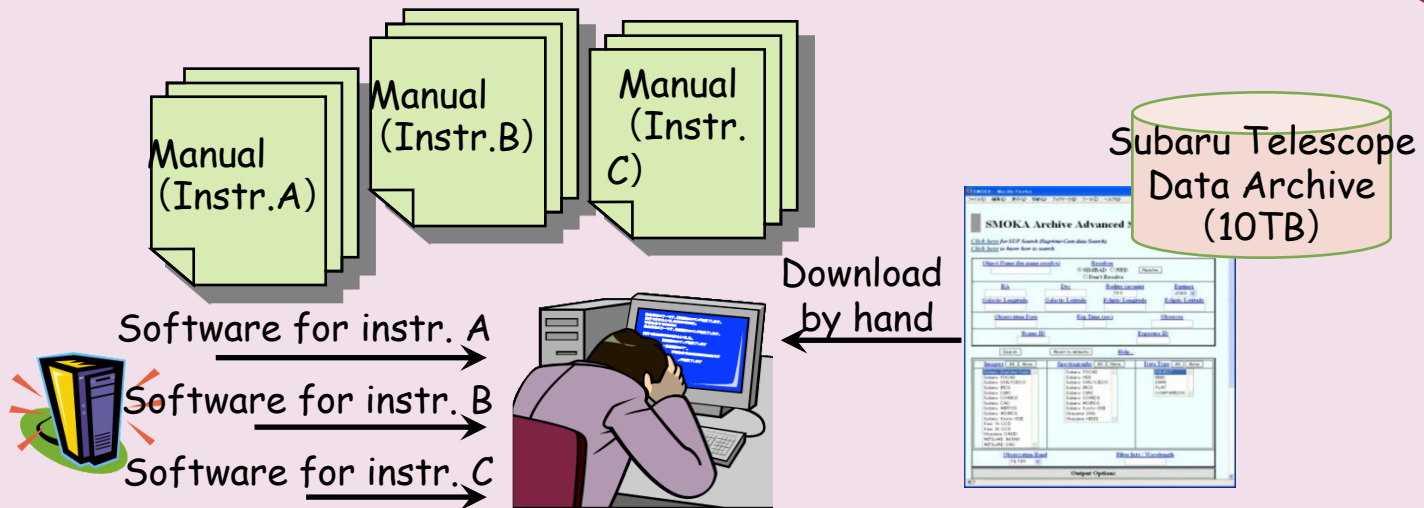


求められる科学成果の最大化

- ▣ すばる望遠鏡 400 億円
 - 445 refereed papers (~2007年) → 1億円/paper
- ▣ TMT 1000 億円？
 - 国内研究者が分担額に見合った成果を上げることが求められる
 - データの占有期間(1年?)内に結果を出さないと...
 - 「どういったサイエンスをするか」に加え、「いかに効率的に成果を達成するか」といった戦略が必要
- ▣ 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。
- ▣ アーカイブデータを利用できる形で配信し、研究成果をさらに促す。

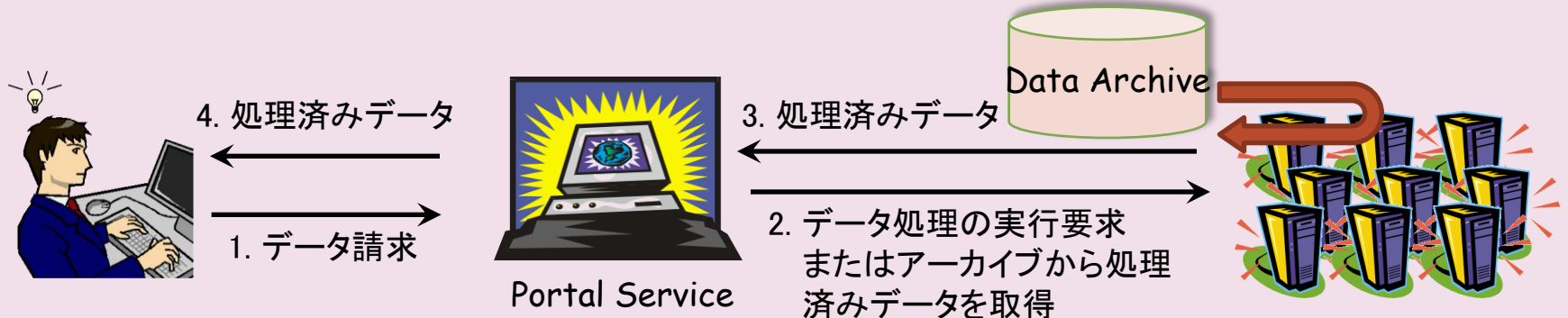
従来型解析手法の限界

- 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
- 正しい方法で解析されていないデータによる間違っただ研究結果が発表されてしまうリスク
- リダクションソフトがすべての OS 環境で正しく動くとは限らない...



リダクション機能つきデータアーカイブ

- データセンターがデータリダクションまでサポート
 - データアーカイブと連携動作するリダクションシステム
 - データ移動のコスト最小化
 - アーカイブ側に計算機資源を集中(並列計算システム)。計算資源の共有。
 - リダクションソフトの管理を一元化。過去バージョンのソフトによる再リダクションをサポート。
 - 請求の多い処理済みデータはアーカイブに蓄積。同じリダクションを繰り返さない。



JVOにおけるSuprime-Cam データ配信

▣ すばるデータリダクション解析システム

- データアーカイブと並列解析システムを専用ネットワークで高速接続 (128Gbpsバックプレーン)
- 合計48 CPU コアにより過去5年間分のデータを約1週間でモザイク画像まで作成可能
- Web ブラウザ IF でユーザがジョブ実行可能

▣ データセット毎にモザイク処理までされたデータを用意。

- 2072セット中1841セットを処理済み
- seeing で選んで4レベルまでのモザイク画像
 - superfine (<0.8"), fine (<1.0"), good (<1.2"), no-selection
- Astrometric Calibration : 0.32" (peak), 1.3" (99%)
- Photometric Calibration : これから...

Suprime-Cam 専用画面



Suprime-Cam [Help\(J\)](#)

Object Name | Date | Reduction | Job Status | Command Queue

Alphabetic: [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) [0](#)

20

Total Number 49

#	ObjectName	<input type="checkbox"/> W-J-B	<input type="checkbox"/> W-J-V	<input type="checkbox"/> W-C-RC	<input type="checkbox"/> W-C-IC	<input type="checkbox"/> W-S-I+	<input type="checkbox"/> W-S-Z	<input type="checkbox"/> W-J-U	<input type="checkbox"/> W-J-VR	<input type="checkbox"/> W-S-G+	<input type="checkbox"/> W-S-R+	<input type="checkbox"/> W-S-ZB	<input type="checkbox"/> W-S-ZR
1	CL0024	<input type="checkbox"/> 2 (1)	0	<input type="checkbox"/> 11 (1)	0	0	<input type="checkbox"/> 8 (1)	0	0	0	0	0	0
2	CL1315+51	<input type="checkbox"/> 8 (1)	0	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0	0
3	CL1320+70	0	0	<input type="checkbox"/> 6 (1)	0	0	0	0	0	0	0	0	0
4	CL1324	0	0	<input type="checkbox"/> 12 (1)	0	0	<input type="checkbox"/> 25	0	0	0	0	0	0
5	CL1520-R	0	0	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0	0
6	CL1604	0	<input type="checkbox"/> 26 (1)	0	<input type="checkbox"/> 2 (1)	0	<input type="checkbox"/> 15 (1)	0	0	0	0	0	0
7	CL1604_0	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 11 (1)	<input type="checkbox"/> 2 (1)	0	<input type="checkbox"/> 26 (1)	0	0	0	0	0	0
8	CL1604_1	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 4 (1)	<input type="checkbox"/> 4 (1)	0	<input type="checkbox"/> 3 (1)	0	0	0	0	0	0
9	CL1604_2	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 4 (1)	<input type="checkbox"/> 4 (1)	0	0	0	0	0	0	0	0
10	CL1604_3	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 3 (1)	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0
11	CL1604_4	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 15 (1)	<input type="checkbox"/> 4 (1)	0	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0
12	CL1604_5	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 6 (1)	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0
13	CL1604_6	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 4 (1)	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0
14	CL1604_7	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 3 (1)	<input type="checkbox"/> 4 (1)	0	0	0	0	0	0	0	0
15	CL1604_8	<input type="checkbox"/> 1 (1)	0	<input type="checkbox"/> 3 (1)	<input type="checkbox"/> 5 (1)	0	0	0	0	0	0	0	0
16	CLJ1226.9	0	<input type="checkbox"/> 6 (1)	0	0	0	0	0	0	0	0	0	0
17	CLJ1350.8	0	<input type="checkbox"/> 2 (1)	<input type="checkbox"/> 3 (1)	0	0	0	0	0	0	0	0	0
18	COSMOS	<input type="checkbox"/> 111 (1)	<input type="checkbox"/> 120 (1)	0	0	<input type="checkbox"/> 115 (1)	<input type="checkbox"/> 431	0	0	<input type="checkbox"/> 171	<input type="checkbox"/> 84	0	0
19	COSMOS_CALIB1	<input type="checkbox"/> 1 (1)	<input type="checkbox"/> 1	0	0	0	0	<input type="checkbox"/> 1	0	0	0	0	0
20	CVn	0	0	0	0	0	0	0	0	<input type="checkbox"/> 4	0	0	0

簡単なインターフェイスによりモザイク画像が取得可能

Mosaic image for 1883 data sets

< 1.2" for 1749 set

< 1.0" for 1542 set

< 0.8" for 1082 set

Suprime-Cam モザイク画像取得用画面

JVO

[Top](#) | [Search](#) | [VO Services](#) | [Subaru](#) | [Analysis](#) | [Workflow](#) | [JVO Space](#)

[\[Logout\]](#)

JAPANESE VIRTUAL OBSERVATORY

p00 ver.080725 [News](#) | [FAQ\(J\)](#) | [Help\(J\)](#) | [Bugs\(J\)](#)

Yuji Shirasaki

⇒ Location: [Top Page](#) > [Subaru](#) > [SPCam](#) > [Object Info](#) > [Mosaic Info](#)

COSMOS W-S-I+ (080620_214814_g08_94) [Help\(J\)](#)

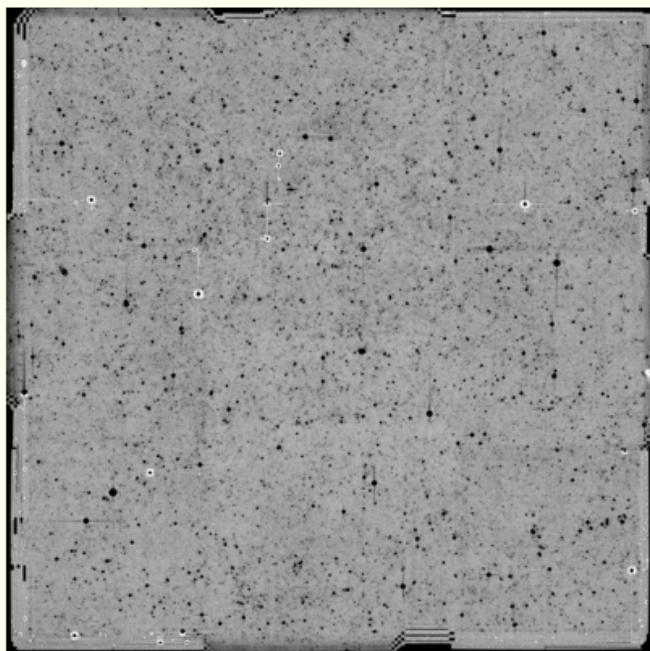
[Summary](#)

[QL Image](#)

[Exposures](#)

[Flat Field](#)

Mosaic Frame ID : SUPM2C3D69E956E0000

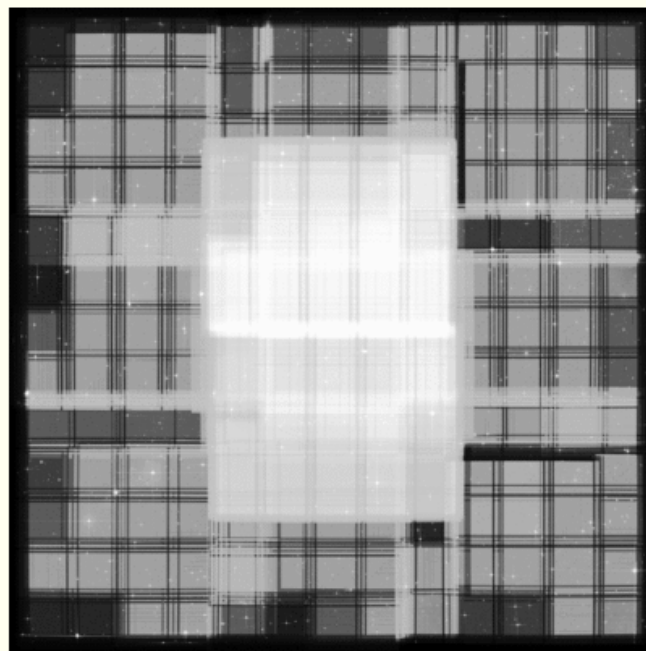


Mosaic frame: [Download](#) 2818.85 MB

Super fine frame: [Download](#) 2798.1 MB (FWHM < 0.8 arcsec)

Fine frame: [Download](#) 2818.84 MB (FWHM < 1.0 arcsec)

Good frame: [Download](#) 2818.84 MB (FWHM < 1.2 arcsec)



Exposure frame: [Download](#) 2818.84 MB

Suprime-Cam モザイク実行用画面

Subaru Suprime-Cam Data JVO Data Viewer JVO Data Viewer

JVO JAPANESE VIRTUAL OBSERVATORY p00 ver.080725 News | FAQ(J) | Help(J) | Bugs(J) [Logout] Yuji Shirasaki

⇒ Location: Top Page > Subaru > SPCam

Suprime-Cam Help(J)

Object Name	Date	Reduction	Job Status	Command Queue
-------------	------	-----------	------------	---------------

mosaic.sh cal-flat.sh command list

RA Dec Size or OBJECT

FILTER

MAX FRAMES MAX humidity (%) MAX seeing (arcsec)

Date (yyyy-mm-dd) From To

Skip Quality Check Only Data Retrieval Skip MOSAIC

Excluded exposures (exposure id, comma separated)

Job Class

Register

OK

完了

- 座標またはオブジェクト名とフィルター名を指定すればモザイク処理パイプラインを実行可能
- 環境パラメータや観測時刻の指定も可能。利用したくないフレームを直接指定することも可能
- 観測装置に固有なパラメータ指定がないので、Suprime-Cam の知識がなくても OK.

TMT Operation Concept Document

▣ 4.4.5 TMT Data Archive

- Discussion: as of December 2006, a TMT data archive is not considered to be part of the baseline TMT construction or operations plans.

...

TMT has two options: build its own data center or form a partnership with an existing data center. Given the large amount of data center expertise that already exists within the TMT Partnership, the latter option seems more attractive and cost effective at this time.

“データベースは既存のデータセンターにお任せしたい”

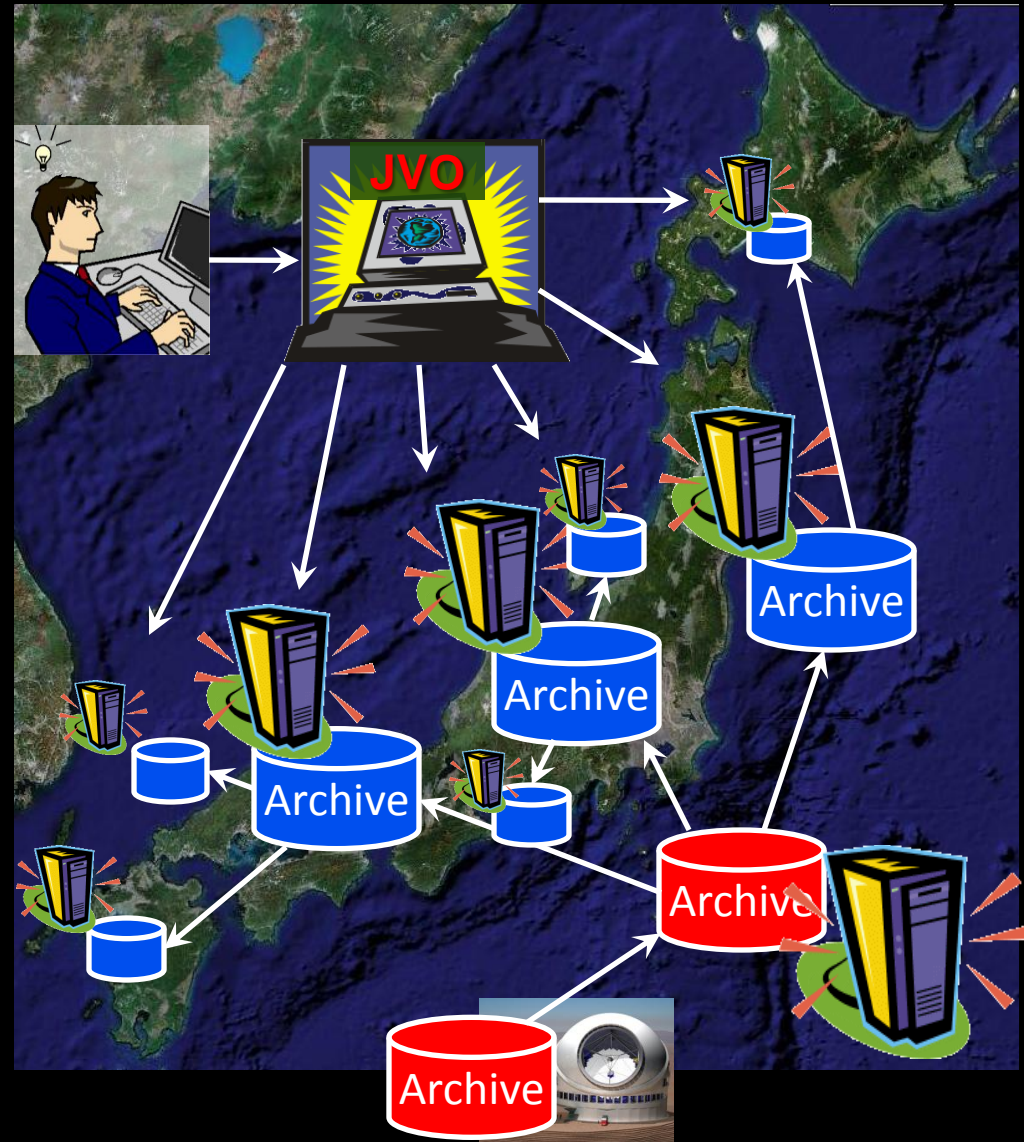
TMT Operation Concept Document

- ▣ 5.2.8 Data processing
 - The TMT Observatory shall take the data processing modules (typically) produced by instrument development teams and create observatory-based data processing pipelines. These pipelines shall be used for quick-look reductions during observing, system performance monitoring, and (where useful) calibrated science products.

“クイックルック用のパイプラインは用意するよ”

国内におけるTMTアーカイブ・解析システム案

- TMT観測所からデータを国立天文台へミラー
- 国立天文台 TMT アーカイブから国内の複数拠点へ再ミラー
- 拠点ミラーからサブセットのみを国内の希望する大学・研究所へ配布
- 各サイトには解析システムも配置。これらの解析システムはデータセンターが全件リダクションを行う際、または一般ユーザがJVO経由によるリダクション実行する場合に利用される。
- 各サイトの解析システムはホスト大学・研究所の研究者は自由に利用可。
- データ移動時間のロスを軽減、計算資源の有効利用。



まとめ

- ▣ 大規模データ時代の到来は目前
- ▣ 大規模データを効率的に処理する計算システムの導入は不可欠
- ▣ データセンターがそうした基盤を整備し、国内研究者が 「大規模かつ高品質データ」で世界と戦える環境を提供していくことが必要
- ▣ そうしたソフトウェア開発を行える人材育成や情報学研究者との連携が重要
- ▣ ADASS 2008 カナダ ケベック 11/2-5
- ▣ ADASS 2009 日本 札幌 10/4-8