

HSC戦略枠サーベイのデータベース

高田唯史

(国立天文台天文データセンター)

山田善彦、小池美知太郎、峯尾聡吾、林裕輔、古澤久徳、池田浩之、安田直樹 他 HSCソフトウェアチーム

何故データベースが必要か？

- データが数(例えば10億程度の天体数、300万ほどのCCDデータファイル数)・量(S15B現在既に圧縮して200TB)共に多すぎて、データそのものを全て持ってくるのは個人的には不可能だし、保管する場所も持ちにくい。
- 自分に必要な分だけ欲しい情報を持ってくるには、多次元空間で定義されたデータ(天体カタログも)だけを高速で拾い出してくる必要がある。
- リレーショナルデータベースを筆頭に、データベースというものは、まさに、多次元空間上で大量な情報の一部分を高速に切り出してくるためにある。

提供されるHSC戦略枠サーベイデータの種類

- 最新のデータリリースはS15Bです。(S14A0、S14A0_b、S15Aもある)
- カタログデータベース
 - サマリーテーブル(1天体について1レコードに情報がまとまっている)
 - Coaddの各バンド画像上の天体カタログ
 - Meas & Forced Source
 - Ccdの画像上の天体カタログ
 - Source & Forced Source
- 画像・カタログファイル(FITSファイル)
 - 処理済みCCD画像(CORR)
 - Warp済みExposure画像(Warp)
 - モザイク補正後のCCD画像(CALEXP)
 - Coaddの各バンド画像(?.?.fitsまたはcalexp)
 - Ccd、Coaddに付属した天体カタログ
 - Match(キャリブ用カタログとのマッチングの結果のカタログ)
 - Src(天体カタログ)
 - Force(天体カタログ)
- 公開用のデータベースのdumpファイル
 - あなたの手元でもデータベース本体を持つことができる、(15TB @ S15B release)

提供されるデータ取得用サービス

- DAS(Data Archive Server)
 - ✓ 画像などのFITSファイル取得用ユーザーインターフェース
 - ✓ 画像ファイル検索(DAS Console)と画像きりだし(DAS Quarry)
- CAS(Catalog Archive Server)
 - ✓ カタログ検索用ユーザーインターフェース
 - ✓ Direct-SQL(直接SQLを打ち込む)とForm(SQL作成サポート用)
- Schema Browser
 - ✓ DBテーブルの中身を記述したもの
 - ✓ 各テーブルに対してカラム名とその中身の記述
- hscMap
 - ✓ 軽量化したCoadd画像を使用したブラウジングツール
 - ✓ 多機能でSQL検索やカタログオーバープロットなどが出来る

データベースソフトはPostgreSQL9.3を使用している

簡易版の検索用ユーザードキュメント:https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/s15b/HOWTO_USE_S15B_DB.html
S15Bデータベースに関する情報:https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/s15b/S15B_database.html

HSC戦略枠サーベイのデータベースの作成方針

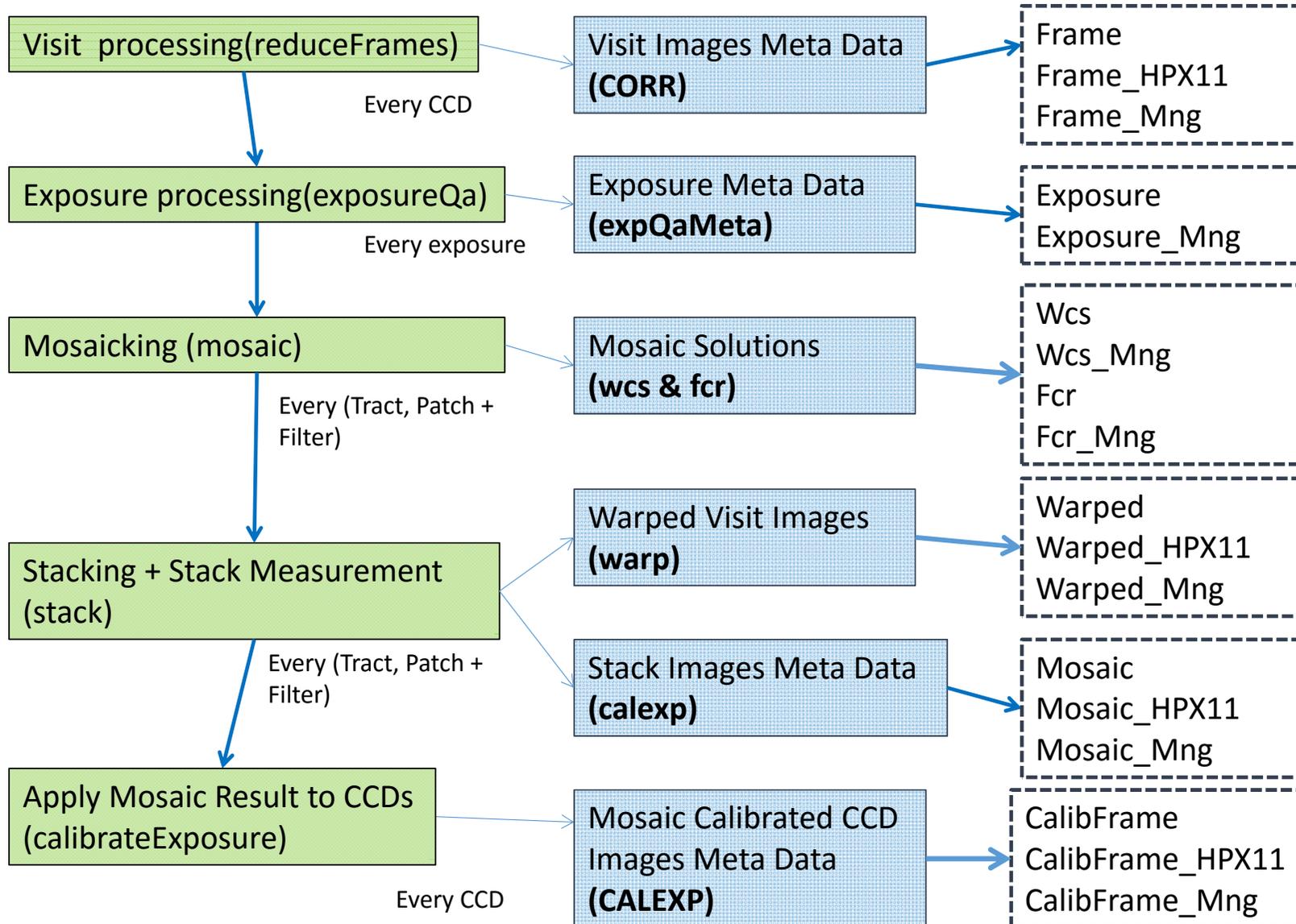
- 画像についてはメタデータ(画像の取得時刻、座標、フィルター、シーイング、測光ゼロ点、イメージクオリティ(点状天体の楕円率)等)をDB登録
- カタログについては基本的に全ての測定情報をDB登録
- パイプラインのはき出す画像とカタログのうち、サイエンスやエンジニアリングに必要なと思われる全ての情報を何らかの形でDBに登録し、検索可能にすることが現状の基本ポリシー

画像のmeta情報DBテーブル

- Meta情報の基本はFITSヘッダー情報(生データ+解析による情報)
- データの種類
 - ✓ 処理済みCCD(CORR)
 - ✓ WCS(モザイクを解いた結果の各CCDの座標に関する変換係数)
 - ✓ FCR(モザイクを解いた結果の各CCDに対するフラックス補正係数)
 - ✓ WARP(各露出の画像をパッチに対して捻って貼り付けた画像)
 - ✓ Coadd(calexp)(複数露出を足し上げた画像)
 - ✓ モザイク情報反映したCCD(CALEXP)

パッチはトラクトとともに画像ファイルの管理単位(詳細は別途)

画像メタ情報



Pipeline Procedures

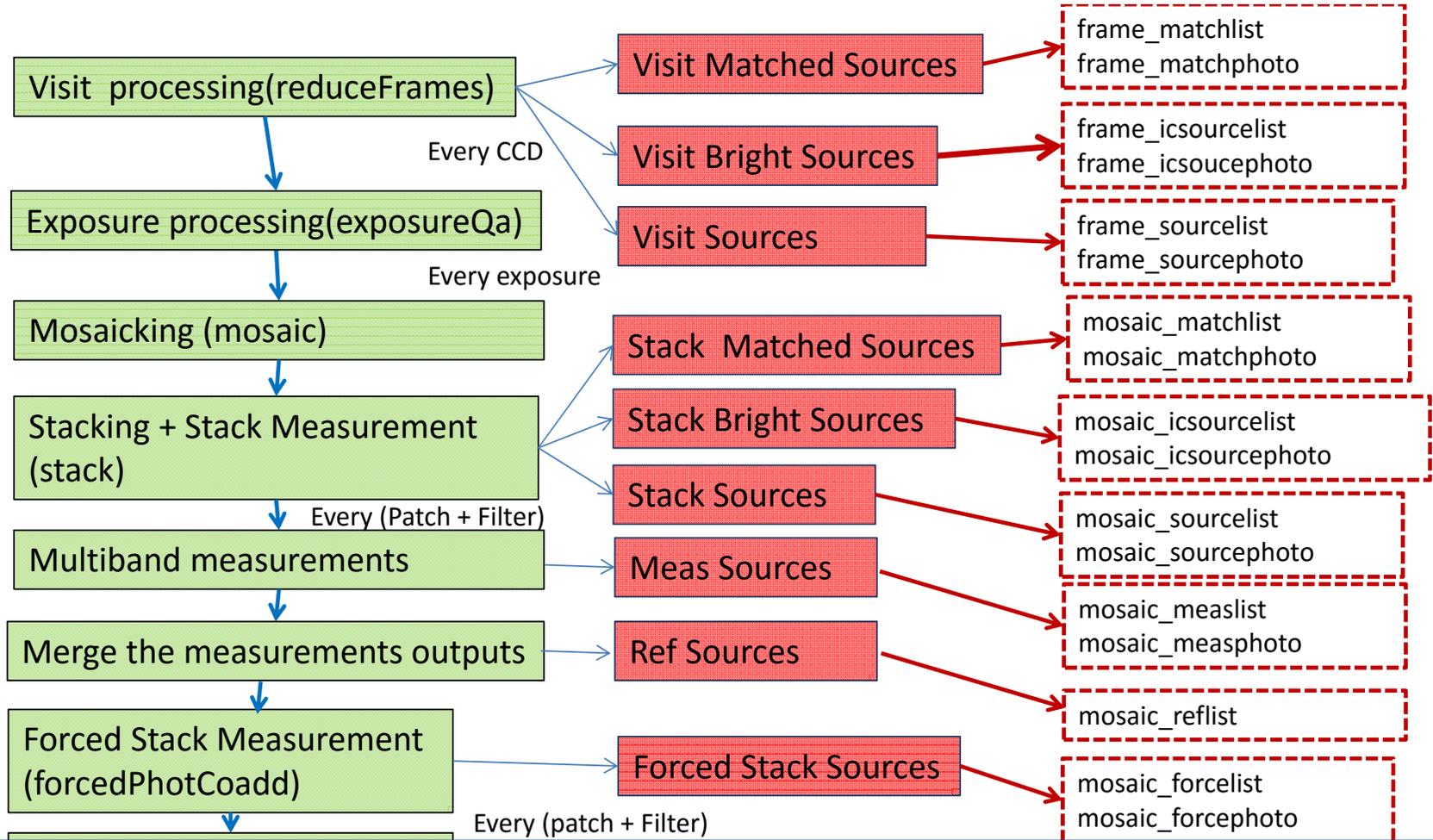
Pipeline Outputs

DataBase tables

カタログDBテーブル

- 解析パイプラインが各画像ファイル毎に書き出した天体カタログをデータベースに入力したもの
- データの種類
 - CCDに関するもの
 - CCD画像に対する検出天体全て(SRC)
 - 検出天体のうち、キャリブレーション用カタログに載っているもの(MATCH)
 - モザイク処理の情報を考慮したCCD画像に対する検出天体全て(CALSRC)
 - Coadd画像で検出された天体についてCCD画像上でforced測光を行ったもの(FORCEDSRC)
 - Coaddに関するもの
 - Coadd画像に対する検出天体全て(meas)
 - 検出天体のうち、キャリブレーション用カタログに載っているもの(meas-match)
 - Coadd画像で検出された天体についてCoadd画像上でforced測光を行ったもの(forced_src)

天体カタログ



Pipeline Procedures

Pipeline Outputs

DB tables

Mosaic_Forcelist + Mosaic_ForcePhoto

ID	Mag1	Mag2	FILTER01
1	23.0	23.1		HSC-G
1	22.9	23.0		HSC-R
1	22.8	22.9		HSC-I
1	22.0	22.1		HSC-Z
1	21.5	21.6		HSC-Y
2
2

サマリーテーブル作成

各バンドのCoadd画像上で検出された同一天体の情報を1レコードに
(Coaddのマルチバンドカタログ: photoobj_mosaic)

ID	gMag1	rMag1	iMag1	zMag1	yMag1	gMag2	rMag2	iMag2	zMag2	yMag2
1	23.0	22.9	22.8	22.0	21.5	23.1	23.0	22.9	22.1	21.6	
2						

Coaddに関するもの (Forced測光)
(Measについては現在作成途中)

Frame_Forcelist + Frame_ForcePhoto

ID	Mag1	Mag2	MJD	FILTER01
1	23.0	23.1	53888	HSC-G
1	22.9	23.0	53889	HSC-G
1	22.9	23.1	53890	HSC-G
1	22.9	22.9	53891	HSC-G
1	23.0	23.0	53892	HSC-G
1			HSC-R
1			HSC-R

各バンドのCCD画像上で検出された同一天体の情報を1レコードに
(時系列データの基本部分) (photoobj_frame)

ID	gMag1	rMag1	iMag1	zMag1	yMag1	gMag2	rMag2	iMag2	zMag2	yMag2	MJD
1	{23.0, 22.9, 22.9, 22.9, 23.0}	{23.1, 23.0, 23.1, 22.9, 23.0}	{53888, 53889, 53890, 53891, 53892}
2						

CCDに関するもの (Forced測光)

サマリーカタログ作成時に付加している情報

- 銀河座標値
- (cx,cy,cz)座標値 ((RA,DEC)を3次元座標 (CX,CY,CZ)に変換したもの)
- HEALPixインデックス (order=18)
- 銀河系による吸収値 (Schlegel et al. 1998)
- 統計値 (forcedCcdのカタログをまとめた場合のみ)

フラッグテーブル

- 各カタログテーブル(mosaic_measlist、mosaic_reflist等という名前のテーブル)の中の検出や測定に関する様々なフラッグ情報(Booleanでの記述がほとんど)を集めてDBテーブルにしたもの
- 欲しい天体群をDBから選び出すときに、正しく測定されていると思われるものや重複がないようにするには必須の情報
- 典型的にはサマリーテーブルと結合(Join)して使われる

テーブルの結合(join)

前提: IDは各天体についてどのDBテーブルにおいても共通 (forced測光及びmeas、meas-match、refについて)

ID	gMag1	rMag1	iMag1	zMag1	yMag1	gMag2	rMag2	iMag2	zMag2	yMag2
1	23.0	22.9	22.8	22.0	21.5	23.1	23.0	22.9	22.1	21.6	
2						

ID	flag1	flag2	flag3	flag4
1	False	False	True	False
2

結合！！

ID	gMag1	rMag1	iMag1	zMag1	yMag1	gMag2	rMag2	iMag2	zMag2	yMag2	flag1	flag2	flag3	flag4
1	23.0	22.9	22.8	22.0	21.5	23.1	23.0	22.9	22.1	21.6		False	False	True	False
2										

IDについて2つのテーブルを結合して1つの横に長いテーブルを”仮想的に”作る

ストアド関数 (Stored Function)

またはUDF(User-Defined-Function)

https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/common/stored_functions.html

- データベース上で動く関数群
 - 統計値計算(weighted mean, median, quantile, skewness, kurtosis他)
 - 時間変換 (MJD<->DATETIME)
 - 座標変換 (赤道<->銀河)
 - 単位変換 (等級<->Flux,Jy,cgs)
 - HSC独自の变換 (HSCA????????? <-> visit、ccd)
 - WCSに関する便利な関数
 - 形状計算 (determinant radius他)

HSC戦略枠データ検索・取得用サイト

(使い方については小池、峯尾両氏の講義にて解説)

- https://hscdata.mtk.nao.ac.jp/hsc_ssp/ “:4443”は無しでも良い
- アクセスするにはユーザーアカウントが必要です。(戦略枠Co-Iの証)
- カタログ検索 (CAS: Catalog Archive Server)
 - Direct SQL Search: 細かい注文型検索
<https://hscdata.mtk.nao.ac.jp/datasearch/> (:4443は無しでも良い)
 - Form Search (SQL Generator): おきまりコース型検索
https://hscdata.mtk.nao.ac.jp/datasearch/photoobj_searches/new
 - hscMap (Explorer: カタログの画像へのオーバープロット)
https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/hscMap/
 - Schema Browser: データベーステーブルの中身を知るためのオンライン辞書
https://hscdata.mtk.nao.ac.jp/schema_browser2/
- 画像・ファイル検索 (DAS: Data Archive Server)
 - フォーム検索 (DAS console): https://hscdata.mtk.nao.ac.jp/das_console/
 - 画像検索切り出し (DAS quarry): https://hscdata.mtk.nao.ac.jp/das_quarry/
 - 直接アクセスによるwgetでのファイル取得も可能
- ストアド関数
 - 検索補助用の関数群 https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/common/stored_functions.html

カタログDB検索のイロハ (Coaddカタログ)

- Template的なSQLをFormUIを使って作成し、それを元にしてDirect-SQL検索用にチューンしていく
 - FormUIは非常に典型的なクエリーを作るための補助ツール
 - 痒いところまで手が届く検索が必要な場合にはDirect-SQLのユーザーインターフェースを使うべし
 - スキーマブラウザを見ながら、必要なパラメータはどのテーブルに載っているかを確認、join等を使って天体毎の情報を統合して使う
 - CASのPreview機能をうまく使って、正しいクエリーを投げられているかの様子を見る
- ダウンロードしたカタログ (CSV等のフォーマット) をhscMap等を使って検出元画像にオーバープロットするなどして様子を見る

簡易版の検索用ユーザードキュメント: https://hscdata.mtk.nao.ac.jp/hsc_ssp/dr1/s15b/HOWTO_USE_S15B_DB.html

データ検索時の注意事項

しつこいようですが、、、

- まずリリースノートを読んでデータ利用に対する制限について理解して下さい。
- データ生成の流れを知り、どのようなデータが取得できるのかを理解して下さい
- 検索の対象となるテーブルの中身についてはSchema Browserを利用して下さい
- ストアド関数(データベース上で利用可能な関数)をうまく利用して下さい。
- フラッグの指定を注意深く行って下さい。
 - ✓ 天体は重複して登録されている(場合がある)
 - ✓ 測定は全てに対して成功しているわけではない
 - ✓ 測定値は明るい星等の影響を受けている場合がある
 - ✓ 天体は明るすぎてサチッたりしている場合もある。。。等々
- フラッグなど別々のテーブルに格納されている情報をJoin機能を使って検索条件やに指定したり、検索の出力として利用して下さい。
- hscMap等を使って、思った通りのサンプルが取得できているのかの確認を行って下さい。

最後に

- 今後、天体カタログ検索用のデータベースは少々変わります、、
 - ✓今よりわかりやすいクエリーを打てる
 - ✓今より反応の速いデータベース
- カタログの種類も今後も増えます！
 - ✓各露出に対する天体カタログをまとめたもの(photoobj_frame等)を今後追加して行きます。
 - ✓Photometric redshift、他のバンドの情報の追加など
- 今回の講習で得た知識は必ず大きく役立ちますので、がっかりしないで下さい、、、 というか、今日からDBを使いたい場合には今日の講習の知識しか意味がありません。