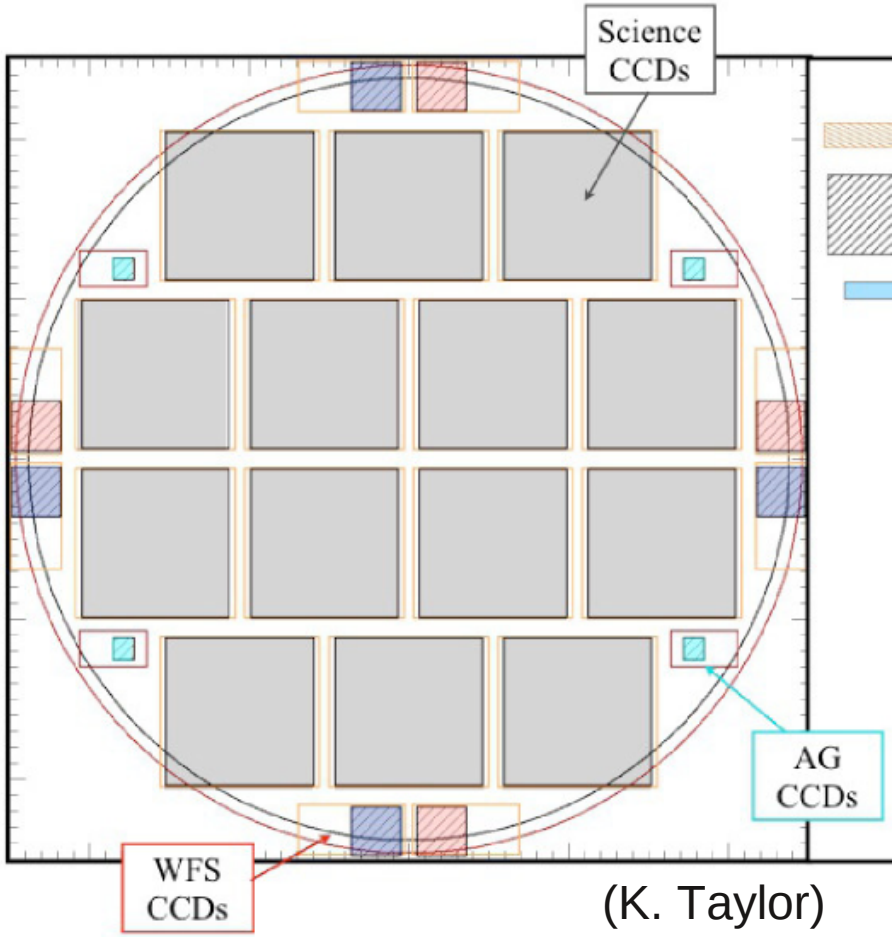
Data services for the J-PAS survey

Paulo Penteado
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Universidade de São Paulo
pp.penteado@gmail.com
http://www.ppenteado.net/ast/pp_ivoa_201210.pdf



(K. Taylor)

Outline

Overview

J-PAS data volume estimates

Comparison with other surveys

Basic Catalogs

Image / Spectrum services

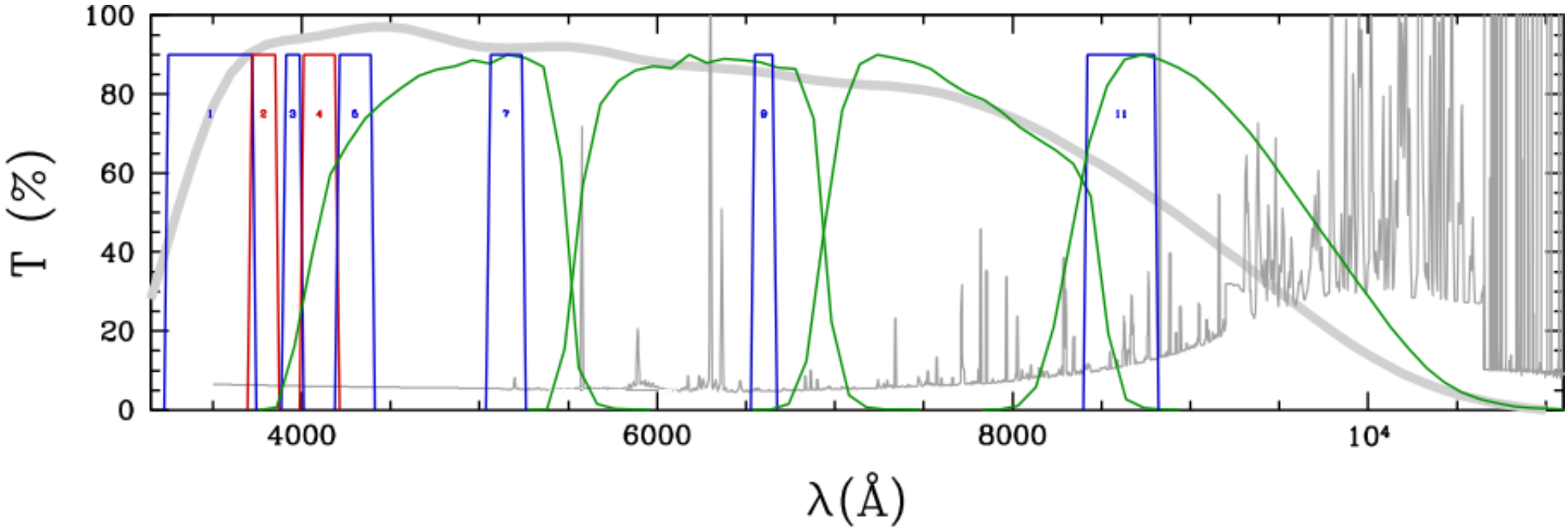
Catalogs of advanced data products

Other possibilities

pp.penteado@gmail.com
http://www.ppenteado.net/ast/pp_ivoa_201210.pdf

J-PAS overview - Filter systems:

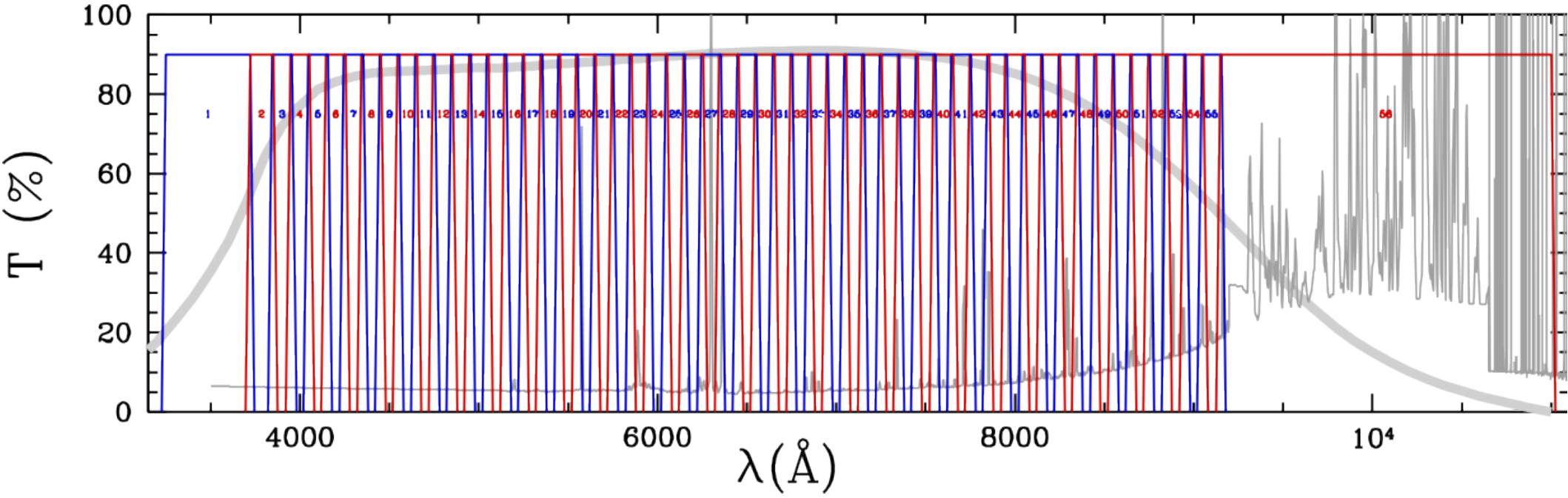
- T80: 12 filters:
  ➔ 4 SDSS (g,r,i,z) + 7:



(A. Marín-Franch)

J-PAS overview - Filter systems:

- T250:
  - ➔ 2 wide-band
  - ➔ 54 narrow-band filters:



(A. Marín-Franch)

J-PAS data volume estimates

- DSS - 1994 - 102 CDROMs - 66.3 GB (images)
- 2MASS - 2006 - 5 DVDs, 43 GB (catalog) 10 TB (images)
- SDSS - 3.4 TB (catalog) 11.6 TB (images)
- **JPAS T80 - ??? (catalogs) 87 TB (images) (~8 times SDSS)**
- **JPAS T250 - ??? (catalogs) 2.3 PB (images) (~ 210 times SDSS)**

(W. Schoenell)

- LSST: 200 PB

J-PAS Basic catalogs:

- ~200 TB.
- With current resources, ~1 month to SEXtract all images.

(A. Amorim)

Data archiving

Storing the data is maybe not that hard:

With the data set increasing linearly with time, and storage doubling every 1.6 years, the peak storage cost occurs 2.3 years into any survey.

From ***A Letter to the NSF Astronomy Portfolio Review: LSST is Not "Big Data"***, David Schlegel (Lawrence Berkeley National Lab), 2012
http://arxiv.org/pdf/1203.0591v1.pdf

What is hard is making the data accessible and useful.

Example 1: SDSS

- Observations in only 5 colors, catalogs have hundreds of columns.
- Many access interfaces, supporting VO standards and ADQL.

Example 2: NASA PDS

- Main repository of all Solar System exploration missions.
- Limited metadata, derived variables and query capabilities for hyperspectral datasets make mining and exploration difficult.

Comparison with other surveys

SDSS

- $8000^{\circ 2}$
- 4×10^8 primary sources and 3×10^8 secondary sources (Photometric Catalog, DR8)
- 157 fields per record
- Easy web interface
- SQL-queries

2MASS

- All-sky ($41253^{\circ 2}$)
- 5×10^8 sources
- 60 fields per record

J-PAS

- Will cover $8000^{\circ 2}$ of the sky.
- Will observe 10^8 galaxies (10^9 stars) in 54+2+1 filters.

(A. Ederoclite)

Basic catalogs

Accessed through TAP (Table Access Protocol) and web interface with ADQL (Astronomical Data Query Language) and VOPlot support.

1) Time-integrated (all exposures coadded)

- 1a) Extended sources
- 1b) Point sources (no morphology columns)
- 1c) Mobile sources (Solar System, high proper motion stars)

2) Time-resolved (individual exposures)

- 2a) Extended sources
- 2b) Point sources (no morphology columns)
- 2c) Mobile sources (Solar System, high proper motion stars)

Basic catalogs

Example proposed columns for basic catalog sources:

ALPHA_J2000,DELTA_J2000, X IMAGE,Y IMAGE
FLUX_APER (0.8”,1.0”,1.5”,2.0”,3.0”,4.0”,6.0”; TBC),
FLUXERR_APER, FLUX_AUTO, FLUXERR_AUTO,
FLUX_ISO, FLUXERR_ISO, FLUX_PETRO,
FLUXERR_PETRO
KRON _RADIUS, PETRO_RADIUS, R_EFF,
FWHM_WORLD
CLASS_STAR
FLUX_MAX,MU_MAX
BACKGROUND, THRESHOLD
FLAG

Per band and “epoch” (56 bands, ~5 “epochs”):
X_IMAGE, Y_IMAGE
FLUX_APER (1 to 6 arcsecs), FLUXERR_APER, FLUX_PSF,
FLUXERR_PSF, FLUX_AUTO, FLUXERR_AUTO, FLUX_ISO,
FLUXERR_ISO
FWHM_IMAGE, FLAGS (A. Ederoclite)

Image / Spectrum services

SIA (Simple Image Access) Service

- 1) Coadded images (best for general use)
- 2) Individual exposures (time resolution, avoid bad frames)

J-PAS is not just a photometric survey: it will make hyperspectral data

So SSAPs (Simple Spectral Access Protocol) services are needed:

- Spectra accessed by sky position (cone), spectral range and time range.
- Plus some optional fields:
  - **`TARGETNAME`** – to get spectra of sources in the catalogs (point, extended and mobile)
  - **`TARGETCLASS, APERTURE`** – frequently needed query fine-tuning

All services with integrated visualization in interactive web interface.

Catalogs of advanced data products

J-PAS will not produce just photometry and spectra.

Data products to be computed by the pipeline and served in catalogs:

- Galaxy parameters: redshift, morphology, luminosity, spectral classification, etc.
- Stellar parameters: spectral type, temperature, mass, stellar type, etc.
- Variability measurements for supernovae, variable stars, Solar System objects
- Detected galaxy clusters (with varied algorithms)
- Whatever else the science team comes up with

Details of these catalogs largely undetermined at this time.

Other possibilities

How to use new algorithms that require access to all (or a large fraction) of the data?

- Download impossible due to data volume
- Server processing slow due to disk / network bottlenecks

We are exploring systems with distributed storage and processing, through multidimensional read-only arrays (i.e., SciDB).

Dynamic function definition capability very desirable for complex data: Allows interactive exploring of mining ideas.

- Not just theoretical argument: One such system, titanbrowse, allowed us to discover the first tropical lakes on Titan:

Possible tropical lakes on Titan from observations of dark terrain. Griffith, C.; Lora, J.; Turner, J.; **Penteado, P.**; Brown, R.; Tomasko, M.; Doose, L.; See, C. Nature, 486, 7402, p. 237-239. 2012. http://dx.doi.org/doi:10.1038/nature11165

Outline

Overview

J-PAS data volume estimates

Comparison with other surveys

Basic Catalogs

Image / Spectrum services

Catalogs of advanced data products

Other possibilities

pp.penteado@gmail.com
http://www.ppenteado.net/ast/pp_ivoa_201210.pdf