

Big Data at the NASA/IPAC Infrared Science Archive (IRSA)

Steve Groom





IRSA: NASA/IPAC Infrared Science Archive

Enabling research that has not yet been envisioned

IRSA's Charter

* IRSA is chartered to curate the science products of NASA's infrared and submillimeter missions.

IRSA's Holdings

- * All sky images and catalogs covering 20 wavelengths.
- Spitzer, NASA's Infrared Great Observatory
- ***** >1 PB since 2014
- * >100 billion table rows



IRSA



IRSA: NASA/IPAC Infrared Science Archive Enabling research that has not yet been envisioned

3

- IRSA's diverse holdings applicable * across the range of NASA's Astrophysics themes
- Richer content draws new, more * complicated potential use cases
- Holding the bits alone is not * sufficient to enable new science
 - > Archive provides support to understand what the bits mean and which ones to use
 - Archive assists users with \succ logistics of access and analysis









Operational Technology

Large Table Management

- * Large tables, growing regularly, while maintaining performance
- HTM indexing used for spatial searches
- * Bulk queries use matching outside the database

Visualization: Firefly by IPAC

- * Archive-aware interactive web tools
- * Linked images, plots, tables
- Reusable components permit both general and project-tailored data views
- * Cost effective to maintain and develop
- * API supports interoperability

	tlas, B1 ¾x			P)	<u>``</u> ç	> 🖉 👎	9				F	5
			می م		COCO COCO COCO COCO COCO COCO COCO COC	2 1 0 -1 -2 -3 -1 ber band zoon	0 1	2 w2m ick and drag to	a pro-w3mpro select an area. (3	4	5	· ·
2-04	800.090	2000	<u>x.04070</u>	<u>8 062</u>	Result IPAC	Table	column Key					
		I 4 4	39 of 42	(3801 - 390	0 of 4118)					Ø 52		ĉ. 0
/1mpro (mag)	w1sigmpro (mag)	K ◀ w1snr	39 of 42 ► w1rchi2	(3801 - 390) w2mpro (mag)	0 of 4118) w2sigmpro (mag)	w2snr	w2rchi2	w3mpro (mag)	w3sigmpro (maq)	Ø ₹ w3snr	w3rchi2	€ ⑦ ₩4
/1mpro (mag)	w1sigmpro (mag)	ld d w1snr	39 of 42 ▶ w1rchi2	(3801 - 390) w2mpro (mag)	0 of 4118) w2sigmpro (mag)	w2snr	w2rchi2	w3mpro (mag)	w3sigmpro (mag)	⊗ ¶ w3snr	w3rchi2	r, @
/1mpro (mag) 560	w1sigmpro (mag)	I ◀ ◀ w1snr 29.1	39 of 42 ▶ w1rchi2 1.258e+00	(3801 - 390) w2mpro (mag) 15.349	0 of 4118) w2sigmpro (mag) 0.081	w2snr 13.4	w2rchi2 8.684e-01	w3mpro (mag) 11.689	w3sigmpro (mag) 0.17	w3snr 6.4	w3rchi2	© @ w4 (r 9.26
/1mpro (mag) 560 412	w1sigmpro (mag) 0.037 0.037	w1snr 29.1 29.2	39 of 42 ▶ w1rchi2 1.258e+00 1.240e+00	(3801 - 390) w2mpro (mag) 15.349 16.771	0 of 4118) w2sigmpro (mag) 0.081 0.297	w2snr 13.4 3.7	w2rchi2 8.684e-01 9.982e-01	w3mpro (mag) 11.689 12.699	w3sigmpro (mag) 0.17	w3snr 6.4 -3.0	w3rchi2 9.325e-01 9.440e-01	© (1 9.26 9.00
/1mpro (mag) 560 412 178	w1sigmpro (mag) 0.037 0.037 0.037	29.1 29.2 29.2	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00	(3801 - 390 w2mpro (mag) 15.349 16.771 15.194	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085	w2snr 13.4 3.7 12.8	w2rchi2 8.684e-01 9.982e-01 9.046e-01	w3mpro (mag) 11.689 12.699 12.813	w3sigmpro (mag) 0.17 0.46	w3snr 6.4 -3.0 2.4	W3rchi2 9.325e-01 9.440e-01 1.052e+00	(1 9.26 9.00 9.05
/1mpro (mag) 560 412 178 447	w1sigmpro (mag) 0.037 0.037 0.037 0.037	29.1 29.2 29.2 29.2	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.105e+00	(3801 - 3900 (mag) 15.349 16.771 15.194 15.371	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.080	w2snr 13.4 3.7 12.8 13.6	w2rchi2 8.684e-01 9.982e-01 9.046e-01 9.072e-01	w3mpro (mag) 11.689 12.699 12.813 12.763	w3sigmpro (mag) 0.17 0.46	w3snr 6.4 -3.0 2.4 0.4	9.325e-01 9.440e-01 1.052e+00 9.716e-01	() 9.26 9.00 9.05 8.94
/1mpro (mag) 560 412 178 447 347	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037	29.1 29.2 29.2 29.2 29.2 29.3	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.105e+00 1.044e+00	(3801 - 390 w2mpro (mag) 15.349 16.771 15.194 15.371 15.116	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.080 0.073	w2snr 13.4 3.7 12.8 13.6 14.9	w2rchi2 8.684e-01 9.982e-01 9.046e-01 9.072e-01 8.382e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337	w3sigmpro (mag) 0.17 0.46	6.4 -3.0 2.4 0.4 1.3	Image: block with the second	(*) (*)
v1mpro (mag) 560 412 178 447 347 522	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037	29.1 29.2 29.2 29.2 29.2 29.3 29.4	39 of 42 ▶ w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.105e+00 1.044e+00 9.519e-01	(3801 - 390 w2mpro (mag) 15.349 16.771 15.194 15.371 15.116 15.465	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.085 0.080 0.073 0.094	w2snr 13.4 3.7 12.8 13.6 14.9 11.6	w2rchi2 8.684e-01 9.982e-01 9.072e-01 8.382e-01 8.382e-01 8.364e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690	w3sigmpro (mag) 0.17 0.46	6.4 -3.0 2.4 0.4 1.3 -1.0	Image: block with the second	9.26 9.00 9.05 8.94 9.03 9.06
/1mpro (mag) 560 412 178 447 347 522 416	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037	w1snr 29.1 29.2 29.2 29.2 29.2 29.2 29.2 29.3 29.4 29.5	39 of 42 ▶ w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.044e+00 9.519e-01 9.390e-01	(3801 - 390 w2mpro (mag) 15.349 16.771 15.194 15.371 15.116 15.465 15.144	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.080 0.073 0.094 0.068	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9	w2rchi2 8.684e-01 9.982e-01 9.046e-01 9.072e-01 8.382e-01 8.364e-01 8.364e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.337 12.690 12.690	w3sigmpro (mag) 0.17 0.46	w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 2.3	w3rchi2 9.325e-01 9.740e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 1.027e+00	 9.26 9.00 9.05 8.94 9.03 9.06 8.69
/1mpro (mag) 560 412 178 447 347 522 416 402	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	w1snr 29.1 29.2 29.2 29.2 29.3 29.4 29.5	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.05e+00 1.044e+00 9.519e-01 9.390e-01 1.262e+00	(3801 - 3900 (mag) 15.349 16.771 15.194 15.371 15.116 15.465 15.144 15.259	0 of 4118)	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9	w2rchi2 8.684e-01 9.982e-01 9.046e-01 9.072e-01 8.382e-01 8.362e-01 9.455e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690 12.690 12.578	w3sigmpro (mag) 0.17 0.46	w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 2.3 0.7	9.325e-01 9.440e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 1.027e+00 9.722e-01	Control (1) Contro(1) Control (1) Control (1) Control (1) Control
/1mpro (mag) 560 412 178 447 347 522 416 402 454	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	29.1 29.2 29.2 29.2 29.2 29.3 29.4 29.5 29.5 29.5 29.6	39 of 42 ↓ w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.05e+00 1.044e+00 9.519e-01 9.390e-01 1.262e+00 1.073e+00	(3801 - 3900 (mag) 15.349 16.771 15.194 15.371 15.116 15.465 15.144 15.259 15.569	0 of 4118) wZsigmpro (mag) 0.081 0.081 0.080 0.073 0.094 0.068 0.073 0.0094 0.068 0.073 0.100	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9	w2rchi2 8.684e-01 9.982e-01 9.046e-01 9.072e-01 8.382e-01 8.364e-01 8.162e-01 9.455e-01 9.455e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690 12.578 12.594	w3sigmpro (mag) 0.17 0.46 0.477	w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 2.3 0.7 0.4	9.325e-01 9.440e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 1.027e+00 9.722e-01 9.722e-01 9.722e-01	9.26 9.00 9.05 8.94 9.03 9.06 8.69 9.13 9.26
/1mpro (mag) 560 412 178 447 347 522 416 402 454 048	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	I I w1snr 29.1 29.2 29.2 29.3 29.4 29.5 29.5 29.6 29.6	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.054e+00 9.519e-01 9.519e-01 1.262e+00 1.073e+00 2.545e+00	(3801 - 3900 w2mpro (mag) 15.349 16.771 15.194 15.371 15.116 15.465 15.144 15.259 15.569 13.426	0.06 f 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.080 0.073 0.094 0.068 0.073 0.094 0.068 0.073 0.094 0.068 0.073 0.094	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3	w2rchi2 8.684e-01 9.982e-01 9.072e-01 8.362e-01 8.362e-01 8.364e-01 9.4552e-01 9.154e-01 1.673e+00	w3mpro (mag) 11.689 12.699 12.813 12.763 12.377 12.690 12.590 12.578 12.594 12.594	w3sigmpro (mag) 0.17 0.46 0.477	w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 2.3 0.7 0.4 -22.3	w3rchi2 9.325e-01 9.440e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 1.027e+00 9.722e-01 9.469e-01 2.902e+00	9.26 9.00 9.05 8.94 9.03 9.06 8.69 9.13 9.26 8.64
/1mpro (mag) 560 412 178 447 347 522 416 402 454 402 454 048 369	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	I I wisnr 29.1 29.2 29.2 29.2 29.3 29.4 29.5 29.5 29.6 29.6 29.6 29.6 29.6	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.105e+00 1.044e+00 9.5190e-01 1.262e+00 1.073e+00 2.545e+00 8.901e-01	(3801 - 390 w2mpro (mag) 15.349 16.771 15.194 15.571 15.116 15.465 15.144 15.259 15.569 15.529 15.529 15.202	0 of 4118) w2sigmpro (mag) 0.081 0.297 0.085 0.080 0.073 0.094 0.068 0.073 0.094 0.073 0.055 0	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3 16.7	w2rchi2 8.684e-01 9.982e-01 9.046e-01 8.382e-01 8.362e-01 8.362e-01 9.455e-01 9.455e-01 9.455e-01 8.672e-00 8.650e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.690 12.578 12.590 12.578 12.541 12.541 12.541	w3sigmpro (mag) 0.17 0.46 0.477	w3snr 4	9.325e-01 9.440e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 1.027e+00 9.722e-01 9.469e-01 2.902e+00 9.118e-01	Comparison of the second
/1mpro (mag) 560 412 178 447 347 522 416 402 454 402 454 048 369 495	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	I I wisnr 29.1 29.2 29.2 29.2 29.2 29.3 29.4 29.5 29.5 29.6 29.6 29.6 29.6	39 of 42 w1rchi2 1.258e+00 1.258e+00 1.206e+00 1.026e+00 1.054e+00 9.519e-01 1.626e+00 1.073e+00 2.545e+00 8.901e-01 1.003e+00	(3801 - 390) w2mpro (mag) 15.349 15.371 15.194 15.371 15.116 15.465 15.144 15.259 15.549 13.426 15.5020 15.368	0.0 of 4118) w2sigmpro (msg) 0.081 0.297 0.085 0.085 0.086 0.073 0.094 0.068 0.073 0.100 0.038 0.065 0.080	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3 16.7 13.6	w2rchi2 8.684e-01 9.082e-01 9.072e-01 8.382e-01 8.382e-01 8.362e-01 9.455e-01 9.455e-01 9.455e-01 8.650e-01 8.650e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690 12.578 12.690 12.578 12.594 12.413 12.414 12.429	w3sigmpro (mag) 0.17 0.46 0.477	w3snr w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 -2.3 0.7 0.4 -22.3 2.3 -0.6	9.325e-01 9.440e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 9.722e-01 9.722e-01 9.722e-01 9.722e-01 9.722e-01 9.722e-01 9.722e-01 9.722e-01 9.572e+00 9.722e-01	Comparison of the second
/1mpro (mag) 560 412 178 447 347 522 416 402 454 048 369 495 334	w1sigmpro (mag) 0.037	I I wisnr 29.1 29.2 29.2 29.2 29.2 29.3 29.4 29.5 29.5 29.6 29.6 29.6 29.6 29.6 29.6 29.6 29.6 29.8 29.8	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.015e+00 1.016e+00 9.519e-01 9.390e-01 1.073e+00 2.545e+00 8.901e-01 1.003e+00 9.275e-01	(3801 - 390 w2mpro (mag) 15.349 15.194 15.371 15.194 15.371 15.465 15.444 15.259 13.426 15.569 13.426 15.500 15.368	0 of 4118) w2sigmpro (msg) 0.081 0.297 0.085 0.080 0.073 0.068 0.073 0.100 0.038 0.065 0.080 0.070	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3 16.7 13.6 15.5	w2rchi2 8.684e-01 9.982e-01 9.072e-01 8.364e-01 8.162e-01 9.455e-01 9.455e-01 9.455e-01 9.455e-01 8.620e-01 8.620e-01 8.729e-01 1.000e+00	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690 12.594 12.594 12.594 12.594 12.594 12.746 12.594	w3sigmpro (mag) 0.17 0.46 0.477	3.0 2.4 0.4 1.3 0.7 0.4 -22.3 2.3 0.6 0.6	With the second secon	9.26 9.00 9.00 9.00 9.00 9.00 9.00 9.00 8.64 9.03 9.26 8.64 8.86 9.05 9.03
✓1mpro (mag) 560 412 178 447 347 522 416 402 454 048 369 334 313	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037	I I wisnr 29.1 29.2 29.2 29.3 29.4 29.5 29.6 29.6 29.6 29.6 29.8 29.8	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.105e+00 1.044e+00 9.519e-01 1.052e+00 1.252e+00 1.252e+00 8.901e-01 1.003e+00 9.275e-01 1.919e+00	(3801 - 390) w2mpro (mag) 15.349 16.771 15.164 15.371 15.165 15.465 15.144 15.259 15.568 15.420 15.388 14.975 16.339	0.081 0.081 0.081 0.085 0.080 0.073 0.094 0.068 0.073 0.094 0.068 0.073 0.073 0.073 0.073 0.073 0.005 0.080 0.065 0.070 0.0231	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3 16.7 13.6 15.5 4.7	w2rchi2 8.684e-01 9.982e-01 9.042e-01 9.072e-01 8.382e-01 8.382e-01 8.382e-01 9.194e-01 1.673e+00 8.650e-01 8.729e-01 1.000e+00 9.690e-01	w3mpro (mag) 11.689 12.693 12.63 12.337 12.690 12.690 12.690 12.690 12.578 12.594 12.541 12.541 12.746 12.840 12.397 12.346	w3sigmpro (mag) 0.17 0.46 0.477 0.471	3 3 3 3 3 3 3 3 3 3	System S	9.26 9.00 9.05 8.94 9.03 9.06 8.69 9.13 9.26 8.64 8.86 9.05 9.03 9.11
/1mpro (mag) 5600 412 178 447 347 522 416 402 454 048 369 495 334 313 433	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.036 0.036 0.036	↓↓↓↓ w1snr 29.1 29.2 29.2 29.2 29.2 29.2 29.2 29.4 29.5 29.6 29.6 29.6 29.6 29.8 29.8 29.8	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.05e+00 1.05e+00 1.05e+00 1.073e+00 2.545e+00 8.901e-01 1.003e+00 9.258e-01 9.588e-01	(3801 - 390 w2mpre (mag) 15.349 16.771 15.194 15.371 15.116 15.465 15.144 15.369 13.426 15.369 13.426 15.368 14.975 16.358	0.06f 4118)	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 10.9 28.3 16.7 13.6 15.5 4.7 12.3	w2rchi2 8.684e-01 9.092e-01 9.072e-01 8.392e-01 8.362e-01 9.455e-01 9.455e-01 9.455e-01 8.650e-01 8.650e-01 8.650e-01 9.649e-01 9.614e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.377 12.690 12.690 12.578 12.578 12.578 12.578 12.579 12.430 12.437 12.430 12.890 12.397 12.445	w3sigmpro (mag) 0.17 0.46 0.477 0.471	Image: With the second seco	9.325e-01 9.325e-01 9.325e-01 9.716e-01 8.339e-01 9.72e-01 9.459e-01 9.459e-01 9.459e-01 9.9118e-01 8.559e-01 9.011e-01 8.855e-01	9.26 9.00 9.05 8.94 9.03 9.06 8.69 9.13 9.26 8.64 8.86 9.05 9.03 9.03 9.01 9.01
/1mpro (mag) 560 412 178 447 522 416 402 454 402 454 048 369 495 334 313 433 389	w1sigmpro (mag) 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.037 0.036 0.036 0.036	↓↓ ↓ 29.1 29.2 29.2 29.2 29.2 29.3 29.4 29.5 29.5 29.6 29.6 29.6 29.8 29.8 29.8 29.8 29.8 29.8 29.8 29.8	39 of 42 w1rchi2 1.258e+00 1.240e+00 2.303e+00 1.05e+00 1.044e+00 9.390e-01 1.65e+00 1.073e+00 2.545e+00 8.901e-01 1.039e+00 9.275e-01 1.919e+00 9.583e-01 1.668e+00 	(3801 - 390) w2mpro (msg) 15.349 16.771 15.194 15.571 15.116 15.425 15.428 15.428 15.428 15.428	0 of 4118) w2sigmpro (msg) 0.081 0.297 0.085 0.080 0.073 0.058 0.073 0.068 0.073 0.068 0.038 0.080 0.070 0.038 0.080 0.070 0.231 0.088 0.087	w2snr 13.4 3.7 12.8 13.6 14.9 11.6 15.9 14.9 14.9 14.9 10.9 28.3 16.7 13.6 15.5 4.7 12.3 16.1	w2rchi2 8.684e-01 9.982e-01 9.072e-01 8.382e-01 8.382e-01 8.162e-01 9.455e-01 9.184e-01 1.673e+00 8.650e-01 8.729e-01 1.000e+00 9.614e-01 8.956e-01	w3mpro (mag) 11.689 12.699 12.813 12.763 12.337 12.690 12.578 12.590 12.578 12.591 12.541 12.541 12.746 12.397 12.446 12.397 12.446 12.397	w3sigmpro (mag) 0.17 0.46 0.477 0.471	€ € 2 w3snr 6.4 -3.0 2.4 0.4 1.3 -1.0 2.3 0.7 0.4 -22.3 2.3 0.6 0.6 0.6 0.8 0.0 3.4	9.325e-01 9.432e-00 9.716e-01 1.052e+00 9.716e-01 8.339e-01 8.327e-01 9.452e-01 9.452e-01 9.452e-01 9.452e-01 9.459e-01 8.559e-01 9.011e-01 8.855e-01 9.498e-01	9.26 9.00 9.05 8.94 9.03 9.05 8.69 9.13 9.26 8.64 8.86 9.05 9.05 9.05 9.03 9.11 9.17 8.50

5





Operational Technology

Data Analysis at the Archive for manipulation of pixel data

- Image Enhancement: HIRES IRAS, Planck
- * WISE/NEOWISE Custom Coadder

HIRES processing of 857 GHz Planck image of M33



HIRES implements the Maximum Correlation Method algorithm developed for IRAS (Aumann, Fowler & Melnyk 1990), adapted to perform resolution enhancement on Planck data



WISE Coadder tool allows users to create customized image coadds (stacks) of WISE data directly from archive

Caltech



Science Exemplifying Big Data

Large scale reprocessing

- Driven by synergy between an IRSA dataset (WISE) and an external dataset (SDSS)
- Resulting product has potential applications to many questions
- * IRSA presently accepts contributed enhanced products

Logistics of Big Data

- * Required access to all the pixel data
- Movement of all the data presented a major challenge
- Processing all the data required compute resources beyond what the archive could offer



Lang (2014) reprocessed WISE single-exposure images to optimize measurements for extended sources.





Science Exemplifying Big Data

Finding features in the data

- New interest in machine learning
- Starting to unlock the potential of discovering patterns by looking at all the data at once



Beaumont et al. (2014) compare machine learning and citizen science searches for interstellar bubbles in Spitzer images.

8



Morello et al. (in prep) are using machine learning on NASA IR catalogs (Spitzer, 2MASS) to identify Wolf Rayet bubbles that are not detectable at visible wavelengths.





Big Data Implications

- * Data discovery tools will become increasingly important selecting the right data
 - > Data flooding in from many projects
 - Interoperability means IRSA users have access through the VO to a multitude of data sets
 - > How can we help users find the ones they are most interested in?
- * Analysis at the archives will become increasingly important
 - > Deep analysis and custom processing requires general access to all the data
 - > Fixed services won't sustain flexible, growing demand
 - > "Big data means you can't move it all", suggesting **the analysis must move to the data**
- * Archives need resources to meet science analysis needs
 - > Potential for improvements to existing services to support out-of-archive analysis
 - > Need resources to stay abreast of technology landscape, important in shaping solutions
 - > Technology pilot efforts given lower priority in 2015 Archives Senior Review







The Archive Paradox

- * NASA's Astrophysics Archives have focused on curation and online query/access services for science datasets. Resources available for custom processing and analysis are limited.
- Mission proposals primarily focused on generation and distribution of standard products
- * Many users have indicated interest in accessing resources for deeper analysis and mining of those products

Incompatible Realities:							
*	Traditionally, archives provide access to data and some limited, shared resources, while researchers/projects cover their specific analysis resource needs						
*	Big data means you can't move it all, it must be analyzed in-place						
Now paradiam poodod?							

New paradigm needed?

09/29/16

Caltech



Possible approaches

- * Partner with NASA and other potential sponsors on developing a more analysis-friendly model, better supporting in-archive processing
- * Use cloud storage and computing resources as a convenient middle ground for shared purposes, some kind of archiving & analysis "co-location" services
 - > Are archive data volumes compatible with cloud cost models?

IRSA

11





Planning for the Future

Panel question #1:

- * What are the processes for planning for future (5—10 years) capabilities of your service?
- * How and from whom do you gather input for this planning process and where does input typically come from?
- * What new feature(s) do you really want to implement?

- * Project Priorities
 - data-providing missions, e.g. WISE, Spitzer, IRTF, SOFIA have strict schedules for data releases and closeout.
 - > Serving mission data releases take high priority
- * User Priorities
 - Twice yearly meetings of our User Panel result in written reports based on presentations.
 - > Once every few years, we survey online users.
 - > Mission teams provide deep, specialized insight.
 - ADEC, NAVO meet once a month and we hear from our sister archives, giving a picture of community expectations.
- * Archive Senior Review
 - > Typically considers 4-5 year window
 - Have requested Big Data studies, e.g. cloud computing, as investments in the future, but unfunded.

09/29/16

12



Retiring Older Technologies

Keeping operations and maintenance under control

Panel question #2:

- * What feature(s) of your service would you like to stop performing?
- * How do you gather input for making such decisions and where does input typically come from?
- * What is preventing you from stopping?

- Some older data sets rely on mission-specific software tools which require maintenance and updates from time to time. Older visualization tools may stop working under newer OS/ browser/language combinations, requiring rewrites.
- These software refresh activities must be incorporated in schedule between high priority activities and development of new functionality
- * Tool refresh tasks discussed with User Panel
- * Tools can be critical to interpretation of the data



09/29/16

Caltech



Interoperability

Panel Question #3:

- * What steps are you taking to make your data interoperable with allied data sets from other data sites in and out of NASA?
- * How do you find allied data sets and what criteria make data sets candidates for enabling interoperability?

- * APIs allow programmatic access to data and tools
 - > Archives can query each other, e.g. NAVO, ADEC
 - > Desktop tools can query archives, e.g. TopCat, ds9
 - Mission pipelines can query archives, e.g. PTF/Spitzer pipelines can query 2MASS catalog for astrometry
 - > All IRSA datasets available through standard API's
- NASA Astrophysics Virtual Observatory (NAVO) ensures NASA archives adopt VO protocols
 - > Catalogs: SCS, TAP already implemented
 - > Images: SIA already implemented
 - > Spectra: SSA in progress
 - > VOSpace: in progress
 - NAVO activities managed by NASA, but provide access to and from non-NASA services, as well
- In the era of big data, we will need to co-locate some datasets to enable full interoperability
 - IRSA gets input from user panel, other ADEC members, NASA missions/scientists, and community to identify candidate data sets



Caltech



14





* Increasing data volume and diversity are attracting new, more complex use cases

* Archives confronting challenges responding to science-driven resource demands

* Fuller science exploitation of the archives' potential requires rethinking the traditional scope of archive responsibilities

IRSA

15

