

Linking Literature to the Data

MAST: Karen Levay, Joshua Peek, Sarah Weissman, Amanda Marrione, Tom Donaldson

STScI Library: Jill Lagerstrom

Acknowledgement: Julie Steffen, Gus Muench, Ethan Vishniac

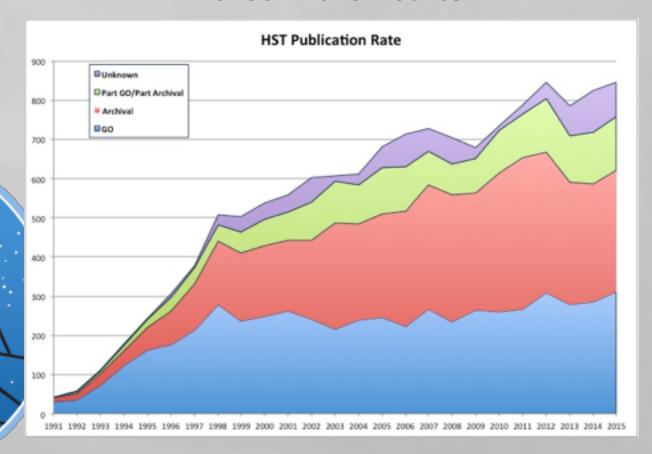


The Goal

- To accurately identify data used in the literature.
- Readers may want to use the same data for their own project.
 - to show how their data have been used.



STScI Bibliometrics





The Challenge

- Authors often have space challenges to describe the data used in their analysis.
- Readers may be challenged to actually identify the specific data used in the paper, especially if the reader is unfamiliar with the relevant instrument.
- Can be time consuming for STScI staff to identify the data.



We searched MAST for HST WFPC2 or WFC3 coeval H α and [O iii] images of PNe available by March 2013. This search yielded H α and [O iii] images for 103 PNe obtained through the F656N and F502N filters, respectively

Galaxy	RA (bhommou)	Dec. (dd:mm:ss)		E(B-V)					Incl. (deg)	Type	
(0)	(2)	(3)	(4)	(5)		(7)				(13)	(14
DDOGGS	10:26:27.78	62:39:25.1	27.82	0.018	15.8		 1.063	0.00	 0.0	1	10

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
DDO078	10:26:27.78	67:39:25.1	27.82	0.018	15.8	-	-	1.063	0.00	-	0.0	I	10.0
IC 4710	18:28:37.95	-66:58:56.1	29.75	0.079	12.51	0.57	11.19	4.494	0.15	-	34.9	Sm	8.9
NGC 1258	3:14:05.50	-21:46:27.3	32.28	0.022	13.88	-	12.35	5.870	0.26	20.5	43.7	SABc	5.7
NGC 3319	10:39:09.47	41:41:12.5	30.7	0.013	11.77	0.41	11.46	7.299	0.51	36.	62.7	SBc	5.9
NGC 5334	13:52:54.44	-1:06:52.4	32.78	0.041	12.97	-	12.19	17,729	0.28	18.2	44.8	Sc	5.2
			100		100	-		100		100		-	-

Notes: The values for all columns are taken from HyperLedu, except for columns 4 and 5, which are taken from NED. More specifically, the distance modulus m — M in column 4 is the median value in NED. If the latter is not available, we adopt the redshift-derived distance modulus, modz, from HyperLedu.

PLG	Contract Value	164	FT 44	Proposal III	PN G	Contract Variety	No.	10.46	Report
		Esp. Time	Esp. Tons				Eqs. Time	Step Stee	
		104	0.0				86	100	
96074 002	K HEH	100	100	100	1967-157	4 6 77	266	266	119
60 (v40)	No. 0.00	1960	200	90%	194 h-454	NAME AND ADDRESS OF	568	1000	
90.71904	B 1 60	400	200	100	100.0	NO THE R. P. LEWIS CO., LANSING	2 miles	1000	
9633-654	80-FF	360	340	100	00.4-30	NOT WHO	440	1000	
MILE PROPERTY.	THE RESERVE	-	100	990	100.0	THE REST	200	1000	\$117, MICH. 95
407-608	W140	440	100	10.00	113-459	86-17	and the last	Section 1	8.0
90.000	WG-89	-	200	100	100.000	10 000	200	200	
40.1-404	MOD WHITE	100	100	1110	94.1-95.1	NOC 1888	100	Date:	
901-8-401-7	90.5-140	100	100	9196	100 1 - 100	MODE STORTS - TO	100	100	
40.4-013	WG-89	346	340	404	MILE - 17 (4)	MOC SIMP	-	-	
90.0-01	B/0, 1	1960	200	9196	2012-042	BC 418		1986	4000, 70
distribution of the last	W 1-0W	340	140	***	200.8-00.0	MACHINE COLUMN	2000	2000	
000.1-010	M(R) 81	360	340	979	224-016	NOT THE	768	100	100
GRAD-5117	Name of Addition	400	160	9179	Distanting.	March Committee	1000	1000	
SER BARRIES	WO-89	400	400	809	580 5-46.0	State of the	366	786	
98.0 (98.	9/913-23	100	(m)	1176	200 Av 50 S	THE RESERVE	100	1.00	\$107, 7800.W
466-1-4013	M 1-00	340	140	809	361 8-46.1	MOCTOR &	100	200	100
SERVICE AND ADDRESS.	WO.49	100	200	1176	275 Avril 12	NAME AND ADDRESS OF	400	1700	400 A
distance of the	963-00	760	140	800	365.4-45.7	Name (medi)		100	
984-99	Way 16-675	100	100	9179	200.01	NO DESCRIPTION	100	1000	
488,0-461	Miles	340	140	100	396.6-06.7	MICH. MICH.	140	104	
BET THREE T	THAT	-	100	100	2011/01/01	Dec 5:00	100.00	1700	EW1.70
distribution in	New Co. DWD		-	***	MILE-1907	MAKES BY	-	1400	-
984-69	Mac 1.11	1960	780	979	100 1-400	NAME AND ADDRESS.	100	200	- 1
distribution?	March 1479	-	***	100	MA horse	March 1989	100	200	
4934-907	MOCHET!	1546	-	480	394-07	SC HEST	540	460	W706.W
204-00	963.0	1040	-	650	594464	PM11-81	-	Distance of the last	160.7
008-018	MODE NOT THE	100	100	100	301.8-10.0	NOT THE	140	1980	
994-91	90140	100	180	8007	2012-02/2	WC III	-	-	
400,0-400	K seed	100	-	1000	86.5-60	Name of Street,	(min	-	400.0
654-03	NOT HER	700	100	WHILE THE RETT	2017-000	No. IT	100	1100	90%
different la	Wilmin	5.00	-	6007	20.0	THE REST	-	CHE	-
C04-1-2	WATER TO	-	140	700, 900	3655-050	NOT WHEN	100	100	- 0
490.00	MOST STATE	100	-	1007	201.0-00.00	MI 1.09	100	100	-
400.71-34(3)	NOT THE	400	180	8114	2012-012	Water St. Std.	366	780	
400 p. dp. 1	200 0 20	-	-	8000	20.6-00.0	81.8	-	-	
90.1-817	MODERN	330	100	470	2013-403	info F	200	200	- 2
9941-011	NOT SEE	100	100		254 Supply	79/04	100	-	
distalled to	Reddom Rehylan	(Mark)	2000	00/9	The Avelland	Thirty and	(m)	-	
073-07	Sec. 2-647	100	-	8007	2014-001	NC 1-14	100	100	
460	THE REAL PROPERTY.	1100	10.00	THE REAL PROPERTY.	FEE 2-100 M	81.9	into	inc.	
90.0-014	NOT WELL	2000	-	8756	796.1-401.7	H1-76	200	700	
90.0-019	Tel 10 (8)	-	100	7600, 6700	200.0-00.00	T 10 P	-	-	
G68 1 - G61 2	W140	_	200	400	790 8-490-4	NC3-08	360	-	
dea hydrod	BOL STREET	-	-	N10. 678	271-007	R1-6	-	-	
900 to 2713	B.F	100	-	470	W13-609	# D- 60	inc.	-	
E24-E2	WINE	100	-	800	79.5-49.7	11-6	-	-	
DITE:	8. in fig.	~	~	-	100.00	Water St. Did.	-	-	
674.5-65.1	NOT SEE		-	100	793-004	E1-01	-	-	
M F	100	-	-	NO. 100	100 June 1	Shirt and	=	-	
90.7+0.3	NOT WELL	-	-	890, 600	291-60	100.7	-	-	- 3
90.5×0.7	200.000								

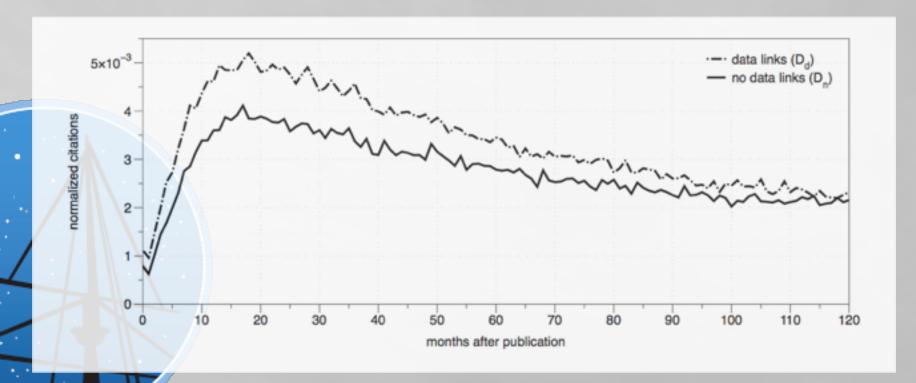
Guerrero+ 2013

Georgiev & Böker 2013

We present a catalogue of photometric and structural properties of 228 nuclear star clusters (NSCs) in nearby late-type disc galaxies. These new measurements are derived from a homogeneous analysis of all suitable Wide Field Planetary Camera 2 (WFPC2) images in the Hubble Space Telescope (HST) archive.

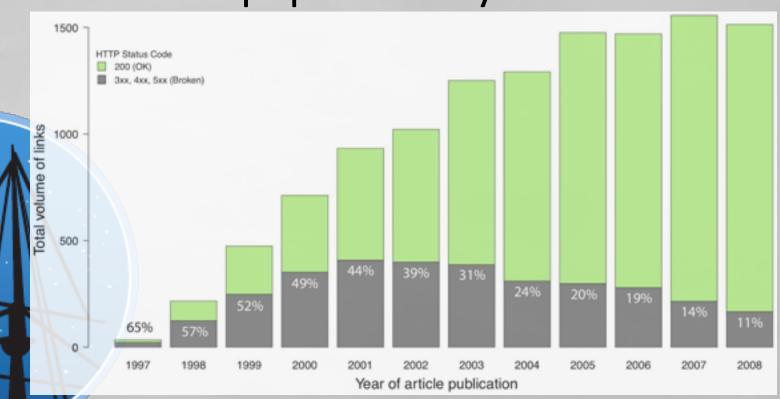


Papers with data links are more highly cited





Links in papers decay over time





MAST and AAS Journals Have Developed a New Workflow

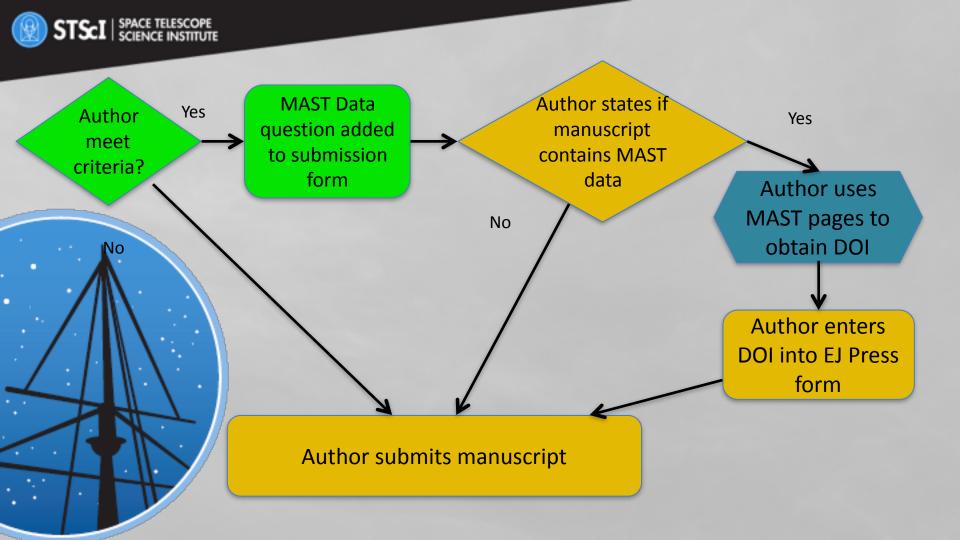
- STScI/MAST can now issue Digital Object Identifiers
 (DOI) or perma-links to define a unique set of data.
- Press will look at each manuscript at submission.
- Author is asked if MAST data is used and if so prompted to submit DOI that can be linked to identify the data used.



Add Datasets

EJ Press submission

MAST (Data hosted at Space Telescope Science Institute) Does your manuscript directly refer to data in MAST (i.e. data from Hubble, Kepler, GALEX, IUE, etc.)? Yes \(\cap \) No What are DOIs for? Data DOIs (permanent links) allow readers to access the data you used directly from the text of your article. How do I get one? MAST provides DOIs for its data in two ways: Find existing DOIs for catalogs (Kepler/KIC, GALEX/MCAT, etc.) and High-Level Science Products (CANDELS, K2SFF, etc.). Generate your own DOI using an interface for making custom collections of observations. Please use this link to the MAST DOI site to find or generate the DOI relevant to your article's MAST datasets. DOI* Clear Clear





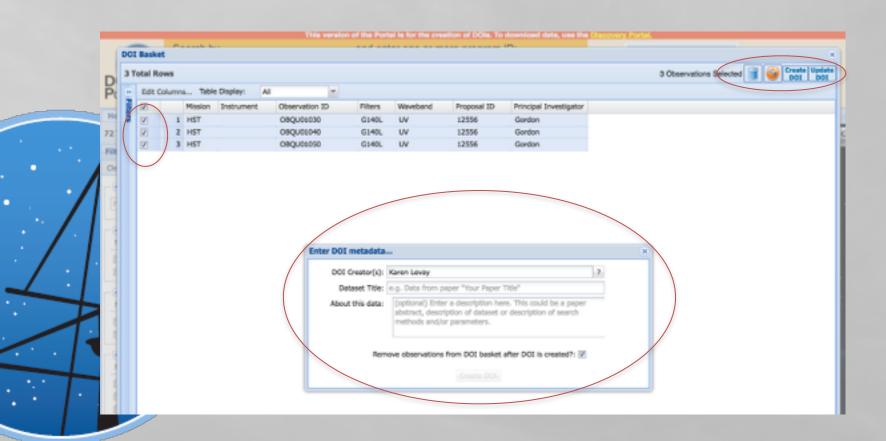
MAST DOI Options

- Some pre-defined DOIs for well defined datasets (e.g. Quarters of Kepler data, GALEX catalog, Community contributed High Level Science Products)
- New interface developed for users to select observations and create a new DOI





DOI Basket







https://archive.stsci.edu/doi/resolve/resolve.html?doi=10.17909/T9RP4V



Test Phase 2016

- First authors from STScI using MAST data through 2016.
- Press will implement process for all papers that appear to use MAST data.
- Plan educational campaign for our users.



Future

- Possible methodologies and feasibility of defining catalog samples from large database (anyone else interested in this topic?)
- Hope to expand our publisher partners.
- Expand use of DOIs for other purposes (e.g. sharing a large data sample among collaborators).
- Improving and expanding "preset" HLSPs.
- Updating existing data links to use DOIs.
- Establish/improve integration with ADS.
- Create DOIs from saved MAST Discovery Portal searches.