

AAS meeting – September 2023



VizieR Staff and contributors:

Astronomers: **P.Ocvirk, G. Monari, C. Bot,**
A. Nebot, S. Derrière

Engineers: **G.Landais, A. Vanhulle,**
T. Boch, F.-X. Pineau

Documentalists: P. Vannier, **E. Perret, C. Fix,**
A. Fiallos, M. Brouty

Outside CDS (engineers): L.Michel,
C. Saillard, T. Keller
(Strasbourg Observatory)



CDS Challenge : quantity vs quality



Preservation commitment :



Data Producers

Space agencies,
Authors, Editors



- Verifications
- Analyses
- Synthesis
- Descriptions
- Standards
- Metadata
- Validations

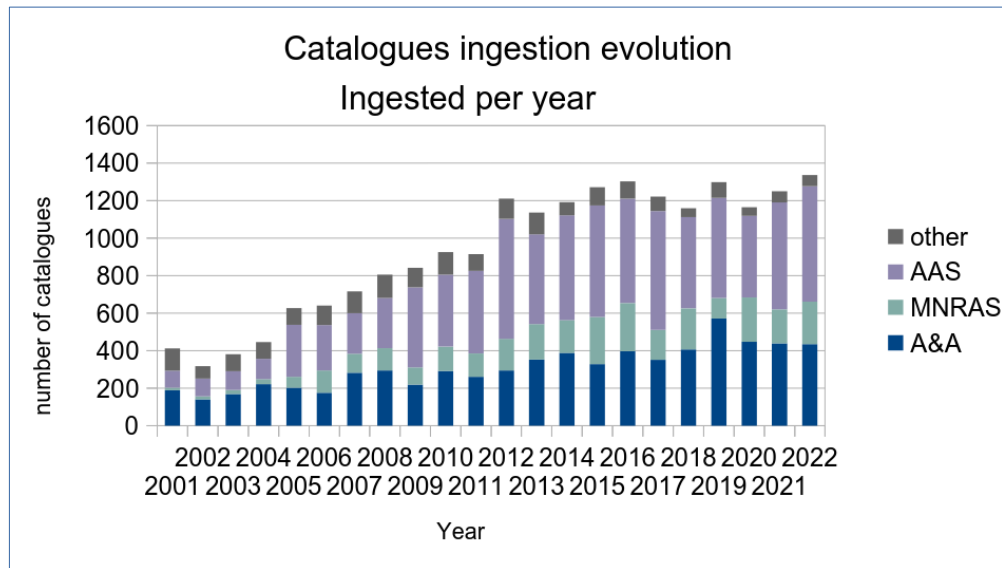
Data Consumers

Astronomers,
softwares (VO),
pipelines

Increasing volume in input vs improving **quality** in output

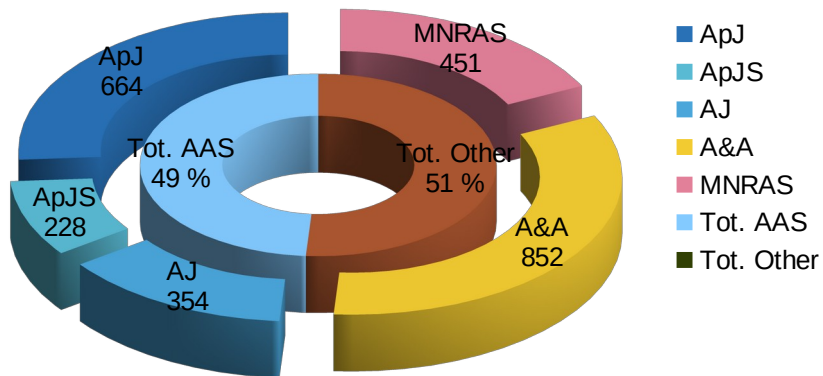


About quantity – AAS is ~half of all ingestions

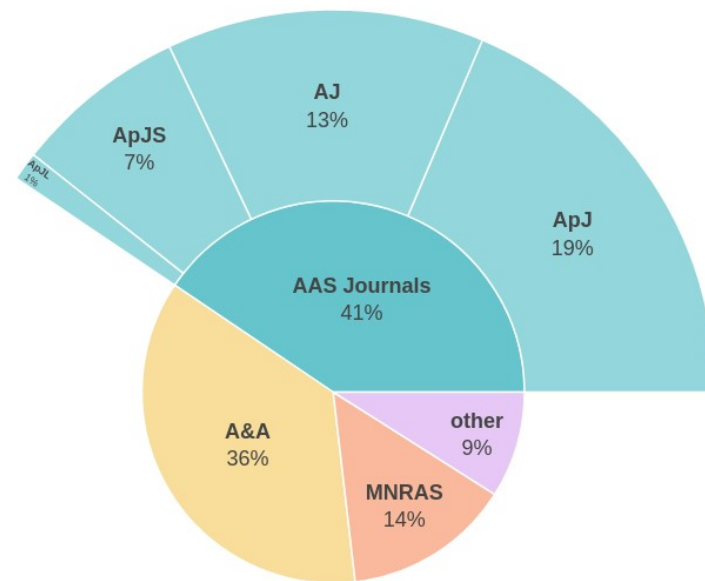


~1250 ingestions per year on average

Number of articles ingested in VizieR 2021+2022



Provenance of the 24168 catalogues in VizieR (2023-09-11)



Currently 41% of the VizieR catalogs are from AAS



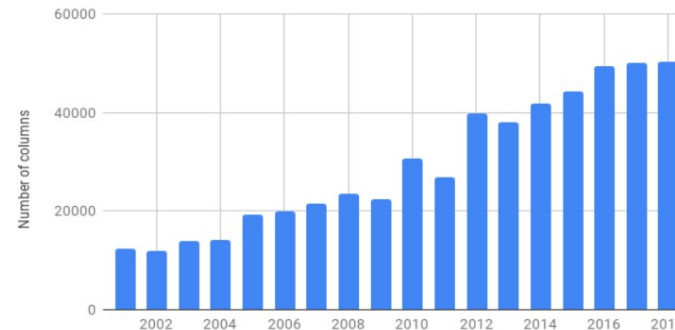
□ Curation challenge : curation evolution



Increasing volume in input

- Number of articles/year published increased slowly
- Number of tables per VizieR catalogue x3 since 2000
- Number of columns per table was ~12.8 in 2000 and ~17 in 2017

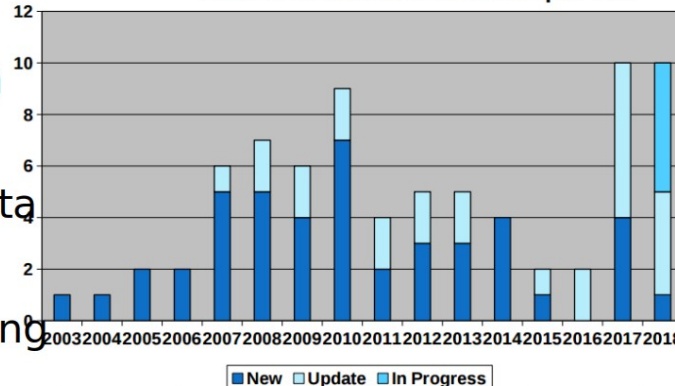
Number of columns evolution per year (S.Derriere)



Evolution and new standards in the VO

- 20 potentials additional metadata to assign
- 10 new tables of metadata among 40 tables

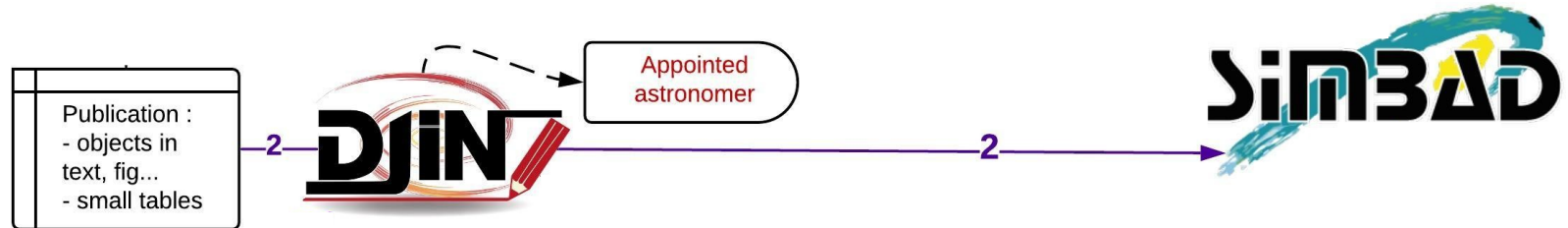
IVOA Standards Recommended per Year



Interop May 2018 – closing session (M.Graham)

The number of publications should not be the only criterion taken into account

Reminder : the workflow – Selection



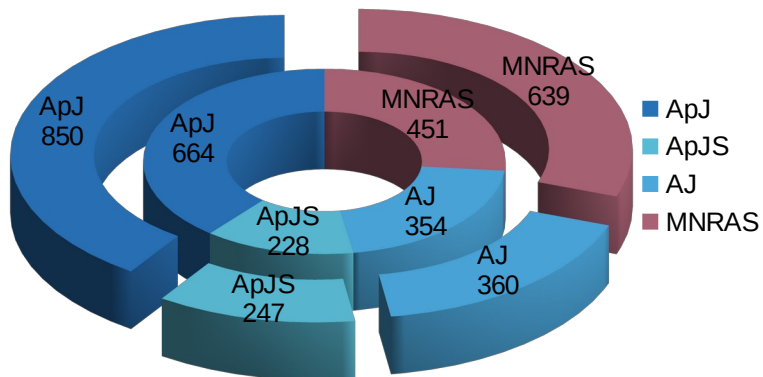
First, Magali (ApJ, ApJS) & Evelyne (AJ) go through **all the articles** to reference astronomical objects from titles, abstract, text, figures, small tables in SIMBAD... ~**No delay** with the publication.

They also warn the VizieR team for « catalogs », « large/complex tables » and « Data behind figures » to process. This is the **main way of selection**.

In rare cases, authors submit directly their data to CDS via the submission tool – currently, this **must be limited to special cases**.



Selection vs Ingestion 2021+2022



- * AJ, ApJS – workload ~absorbed
- * ApJ – backlog => workload excess (currently processing Feb. 2022)
- * Apart from A&A (specific case), AAS is widely ingested

Not all selections become ingestions:

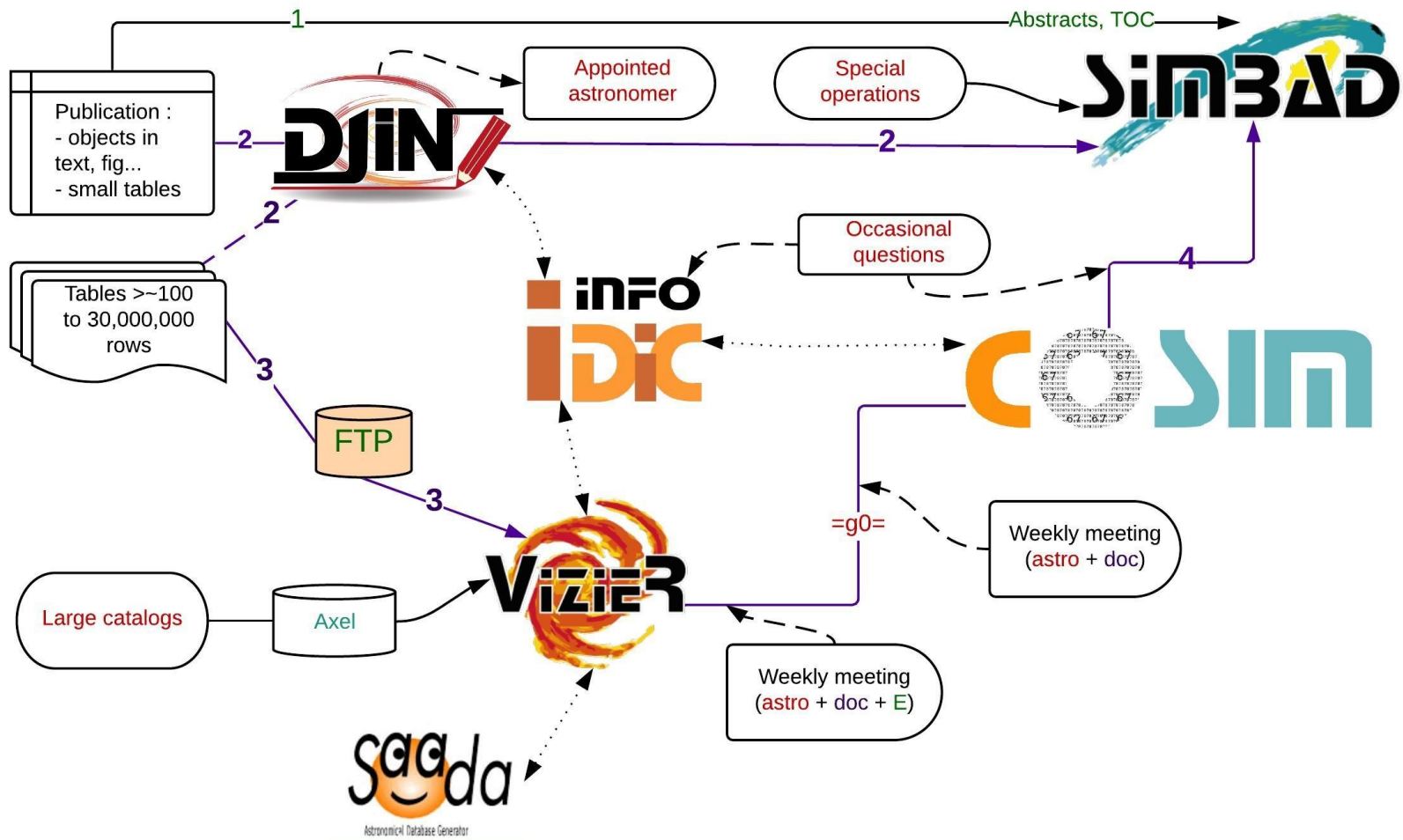
- * priorities for observations or astronomical objects for SIMBAD
- * rejections (too small MRTs, no interest for VizieR)
- * non-recoverable data

Reminder : the workflow

- VizieR is one step (not linear) of the CDS chain



Diverse interactions between documentalists, astronomers and engineers



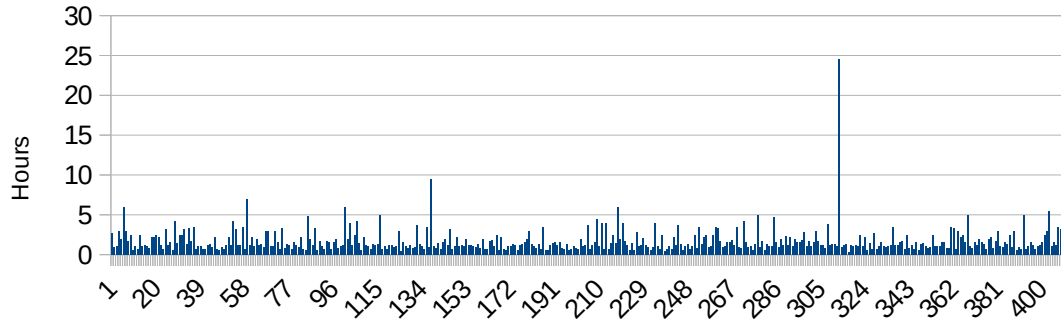


Often asked question: the time spent per cat.? (for both quantity & quality)



Statistics on a subsample for which we have the time spent per catalog :
ApJ & ApJS ingested in the years 2021+2022 by one documentalist

2021+2022 ApJ hours per article

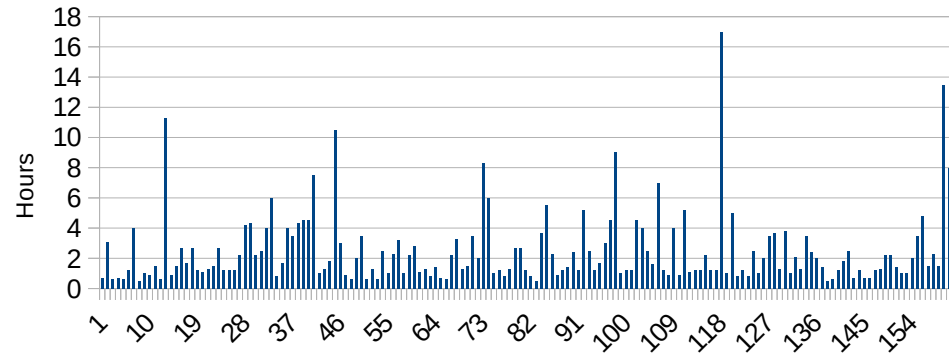


Very variable

Between 0.3h and 24.5h per catalog for ApJ+ApJS on years 2021+2022.

On average, 1.7h for ApJ vs 2.5h for ApJS per catalog on years 2021+2022.

2021+2022 ApJS hours per article





About volume increase solutions



MAKE
YOUR DATA
VISIBLE

VizieR

- **Greater data selection:** priority to observations
– but requirements for SIMBAD too.
- **One additional documentalist hired** since 2013
– but temporary contract renewed every ~3-4 years, whereas training lasts ~6 months.
- **Use as many tools as possible to automate** what can be automated (colmeta, getapj, new setUCD...)
– but engineers are in short supply for VizieR
- **Encouraging authors to follow the Best practices**
– sometimes works...

On the publication support page



« Make your data visible » – EWASS 2019

Raise the odds to be included in VizieR!

Why a checklist?

What about FITS spectra/images?

Ensure that your published tables will be usable by following this **checklist** of essential points:

To face the increasing amount of data, the CDS selects the articles to be processed with various scientific criteria but also by the effort required to make the data reusable.

VizieR offers a specific search for images, spectra and data-cubes in FITS format.

❑ For tables containing known astronomical objects, an existing **non-altered name** must be given along with the **coordinates**.

A table of astronomical objects with no coordinates will not be discoverable in VizieR, and will not be processed in SIMBAD. Shortened names like 2M1427+3400 can be ambiguous because they can be matched to multiple objects; use the full **non-altered** object name instead:
2MASS J14270471+3400138

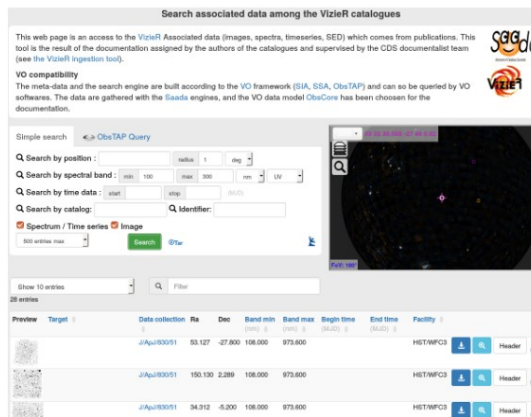
❑ All columns must be well explained, with **their corresponding unit**.

If too many conditions are needed to understand a column, an additional script is needed to process it, and so, it is not easily interoperable.

Please make your columns homogeneous, and avoid mixing measurements with different meanings: errors mixed with limits, or values with different units (that should be in different columns).

❑ When there is more than one table, the objects in common must be identified with the **same name between tables**.

A common key is fundamental for any action between tables.



A good FITS header is the key for reusability - the CDS encourages the usage of the FITS standards: <https://fits.gsfc.nasa.gov/>

Check that FITS file headers include:

- coordinates and wavelengths in the WCS system.
- observation dates.
- telescope/instrument names.

If the FITS file is not in the standard WCS system, the reusability of the file is compromised.

Try this tool to check the completeness of your FITS header:
<http://cdsarc.u-strasbg.fr/vizier.submit/fitsvalidator.html>



Only 3 points...

But most time-consuming



New time-saving programs (not perfect but...)



colmeta : Help to format the ReadMe file – especially helpful to describe coordinates & usual parameters from one standard label and follow the rule to 80cc-limited...

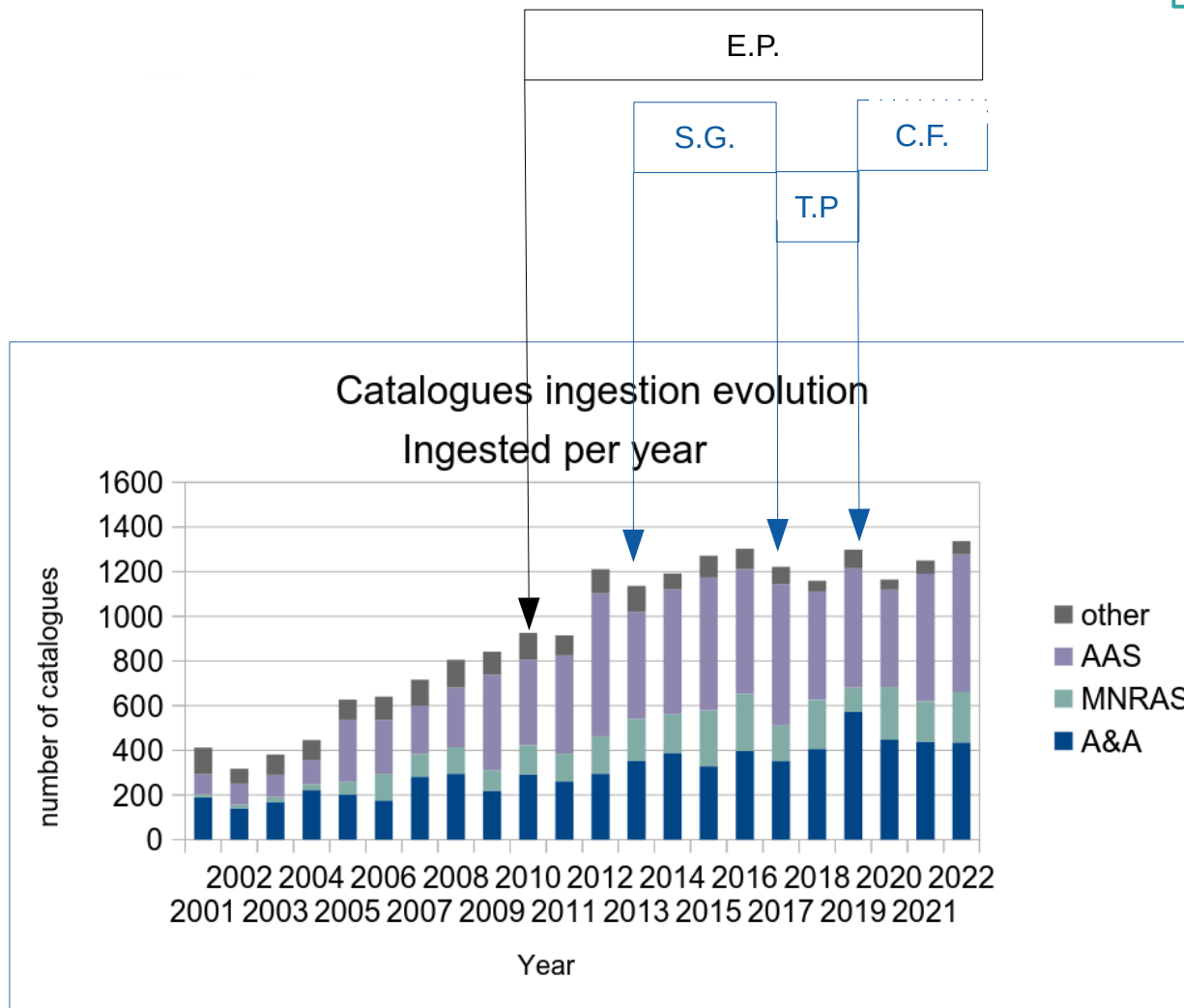
getapj : uses the BCS to retrieve **MRT tables** and create a ReadMe file from them – especially helpful for multi-tables in one catalog + the « See also » section and min/max values are automatically added.

setUCD : the new tool should improve the assignation of ucd1+



New documentalists hired

This is obviously not the only reason for the variations in the quantity ingested per year (it depends in particular on the time spent by the team on ingestions alone, in relation to their other activities, pandemic crisis and so on) but it seems logical to see a **drop when recruiting**, due to the time spent on training, **before it picks up again**





Workflow insights – easy case – Before/After :2021ApJ...913..143G ~30 minutes

Actually, not ~30min. straight :

* first step: creating the ReadMe file and standardize the tables; put in place the commands for VizieR, first set of verifications (~3/4 of the work for an easy-case).

* final step is **few weeks later (normally ~1 month)**: re-read the ReadMe file and ingest the catalog

Title: The Luminous and Double-Peaked Type IC Supernova 2019stc: Evidence for Multiple Energy Sources
 Authors: Gomez S., Berger E., Hosseinzadeh G., Blanchard P.K., Nicholl M., Villar V.A.
 Table: Data behind Figure 2 -- Optical light curves of SN 2019stc in the griz bands
 =====
 Byte-by-byte Description of file: dbf2.txt

Bytes	Format	Units	Label	Explanations
1- 14	F14.8	d	MJD	Modified Julian Date of observation, JD-2400000.5
16- 23	F8.5	mag	omag	Observed AB magnitude in Filter
25- 33	F9.6	mag	e_omag	?=-1.00000 Uncertainty in omag
35- 35	A1	---	Filter	Filter
37- 45	A9	---	Tele	Telescope/Instrument used (1)
47- 51	A5	---	upperlimit	Is the measurement an upper limit (e_omag=-1)
53- 54	A2	---	System	Photometric System

 Note (1): Telescope/Instrument used --
 ZTF = Zwicky Transient Facility, Palomar 48-inch Telescope;
 KeplerCam = KeplerCam Imager, Fred Lawrence Whipple Observatory 48-inch Telescope;
 Binospec = Binospec Spectrograph, MMT Telescope;
 LDSS3c = Low Dispersion Survey Spectrograph, Magellan Telescopes;

58745.49609375	21.35643	0.231554	r ZTF	False AB
58750.50390625	20.31061	0.075235	r ZTF	False AB
58756.51953125	19.97693	0.068371	r ZTF	False AB
58763.50781250	19.34048	0.145183	r ZTF	False AB
58766.50390625	19.22941	0.139155	r ZTF	False AB
58769.46484375	19.12832	0.054426	r ZTF	False AB
58772.53906250	18.87869	0.056052	r ZTF	False AB
58778.48437500	18.82999	0.046348	r ZTF	False AB
58781.49609375	18.94375	0.060983	r ZTF	False AB
58784.51953125	18.77436	0.084771	r ZTF	False AB
58787.52343750	18.83211	0.064434	r ZTF	False AB
58793.50781250	18.69459	0.051258	r ZTF	False AB
58796.53906250	18.74523	0.129825	r ZTF	False AB
58799.36328125	18.74872	0.035149	r ZTF	False AB
58806.49609375	18.77615	0.118572	r ZTF	False AB
58812.30859375	18.96452	0.059139	r ZTF	False AB
58827.28515625	19.25720	0.070847	r ZTF	False AB
58833.37890625	19.44682	0.070968	r ZTF	False AB

Data behind Figure 2 – original data (MRT)



Workflow insights – easy case – Before/After :2021ApJ...913..143G ~30 minutes

Byte-by-byte Description of file: dbf2.txt

Bytes	Format	Units	Label	Explanations
1- 14	F14.8	d	MJD	Modified Julian Date of observation, JD-2400000.5
16- 23	F8.5	mag	omag	Observed AB magnitude in Filter
25- 33	F9.6	mag	e_omag	?=-1.00000 Uncertainty in omag
35- 35	A1	---	Filter	Filter
37- 45	A9	---	Tele	Telescope/Instrument used (1)
47- 51	A5	---	upperlimit	Is the measurement an upper limit (e_omag=-1)
53- 54	A2	---	System	Photometric System

Note (1): Telescope/Instrument used --
 ZTF = Zwicky Transient Facility, Palomar 48-inch Telescope;
 KeplerCam = KeplerCam Imager, Fred Lawrence Whipple Observatory 48-inch Telescope;
 Binospec = Binospec Spectrograph, MMT Telescope;
 LDSS3c = Low Dispersion Survey Spectrograph, Magellan Telescopes

Byte-by-byte Description of file: fig2.dat

Bytes	Format	Units	Label	Explanations
1- 14	F14.8	d	MJD	[58745.49/59174.5] Modified Julian Date of observation (JD-2400000.5)
16	A1	---	l_omag	Upper limit flag on omag
17- 24	F8.5	mag	omag	[18.69/25.77] Observed AB magnitude in Filter
26- 34	F9.6	mag	e_omag	[0.02/0.6]?=-1 Uncertainty in omag
36	A1	---	Filt	Filter (g, r, i or z)
38- 46	A9	---	Tel	Telescope/Instrument used (1)
48- 52	A5	---	uLim	Is the measurement an upper limit; "False" or "True"; e_omag=-1 if "True"
54- 55	A2	---	Syst	Photometric System (always "AB")

Note (1): Telescope/Instrument used as follows:
 ZTF = Zwicky Transient Facility, Palomar 48-inch Telescope (66 occurrences);
 KeplerCam = KeplerCam Imager, Fred Lawrence Whipple Observatory 48-inch Telescope (63 occurrences);
 Binospec = Binospec Spectrograph, MMT Telescope (15 occurrences);
 LDSS3c = Low Dispersion Survey Spectrograph, Magellan Telescopes (9 occurrences);

Original data

Description:

We obtained optical images of SN 2019stc in the griz filters with three different telescopes: KeplerCam on the 1.2m telescope at Fred Lawrence Whipple Observatory (FLWO), the Low Dispersion Survey Spectrograph (LDSS3c) on the Magellan Clay 6.5m telescope at Las Campanas Observatory, and Binospec on the MMT 6.5-m telescope.

We include additional photometry from the Zwicky Transient Facility (ZTF) images. We downloaded the original ZTF images from the NASA/IPAC Infrared Science Archive.

We also obtained seven epochs (spanning 2019-Nov-23 to 2020-Nov-19) of low-resolution optical spectroscopy covering phases from 15 to 340 days. We used the LDSS3c Spectrograph and Inamori-Magellan Areal Camera and Spectrograph (IMACS) on the Magellan 6.5m telescopes and the Blue Channel and Binospec spectrographs on the MMT 6.5m telescope. See Section 2.3.

Objects:

RA	(ICRS)	DE	Designation(s)
06 54 23.10	+17 29 31.3		ZTF19acbonaa = SN 2019stc

File Summary:

FileName	Lrecl	Records	Explanations
ReadMe	80	.	This file
fig2.dat	55	153	Optical light curves of SN 2019stc in the griz bands

See also:

- J/ApJ/741/97 : Light curves of Ibc supernovae (Drout+, 2011)
- J/AJ/147/99 : Spectroscopy of 73 stripped core-collapse SNe (Modjaz+, 2014)
- J/A+A/618/A37 : Spectral log of (i)PTF stripped-envelope SN (Fremming+, 2018)
- J/ApJ/881/87 : 9 epochs spectroscopy of type I SN 201iet (Gomez+, 2019)
- J/ApJ/887/169 : UV-Opt light curves of the type Ic SN 2018gep (Ho+, 2019)
- J/ApJ/901/61 : LCs of 4 superluminous SNe from the ZTF survey (Lunnan+, 2020)
- J/ApJ/902/L8 : Optical & NIR spectra of ZTF19aawfbtg (SN2019hge) (Yan+, 2020)

ry:
 rom electronic version of the journal

ReadMe



Workflow insights – difficult cases

FITS / ASCII & other file formats
 => Tables need to be transformed to the **standard** aligned ASCII file format

For FITS :

- format (including significant digits) for each parameter is never given

=> each column must be filled with a format by hand

- vectors are not supported by VizieR
 => duplication of X columns per value in the vector (same format).

What helps :

CSV format ; the best is the **MRT format** which includes units & explanations in addition to the formats

Top 10+10 of most consuming-time catalogs in 2021-2022 (for those having time record)

ApJ		ApJS	
Total hours	Catalog	Total hours	Catalog
24,5	RELICS, 41 catalogs ASCII (from MAST)	17	SDSS RM, FITS with 276 col. Including vectors
9,5	3 cat. ASCII From author's archive + 9258 LC	13,5	10 tables in ASCII with bib. references, Abbreviated names+comp. not always the same between tables Not always coo.
7	VLA FF survey 14 tables ASCII (inc. from online archive) + Images FITS	11,3	SDSS-IV MaNGA FITS with 535 columns
6	3 tables MRTs + 2 behind Fig. +1 ASCII added for links toward spectra One name modified (exchange with author)	10,5	7 tables in ASCII Script to add coo in a table from another table of the same paper (not clear otherwise)
6	4 tables MRTs (T1 ~220 col.) + Table of 846 references	9	The XMM-SERVS survey 3 MRTs (inc. 206 & 197 col.)
6	4 tables (1 MRT, 3 ASCII) Curation of T2 complicated : wrong coo ; different values for NULL. 145 sources without RA (author has never answered)	8,3	IDEOS 1 table MRT , 232 col. SimbadName for 3559 sources
5,5	3 small Tables in MRTs Issues to link tables + SED Data presented in another paper + erratum (2 tables to add – via script; paste impossible)	8	22 tables en MRTs Names between tables for the same objects are not identical & not well written => misprints/comp. missing (diverse exchanges with the author)
5	1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII	7,5	BASS XXII. BASS DR2 AGN cat. 4 other papers of the same series to do. Flags added to link other tables let in the paper 2 MRTs tables + 1 ASCII Retrieving names from SWIFT.
5	Cosmicflows-4 2 MRTs (inc. 85 col). No coo. Retrieved from LEDA, 2MASX... => merge/lds added in SIMBAD + rewriting names for ~300 galaxies	7	BDKP. Paper V. 7 tables (3 MRTs, 4 ASCII) Names between tables => misprints – diverse exchanges with the author
5	1 MRT (~90 col.) No coo. 2 col. Simbad for 6393 pl. + stars	6	8 tables (6 en MRTs + tables added: notes & abund) SimbadName



Workflow insights – difficult cases

Number of tables per cat. increase

=> links between tables become more complicated

The links help to find misprint/missing data; added-value: number of galaxies per clusters...

What helps:

- Keep a same ID between tables!
See « Make your data visible », 3rd point
- Keep the same format between tables.

ApJ		ApJS	
Total hours	Catalog	Total hours	Catalog
24,5	RELICS, 41 catalogs ASCII (from MAST)	17	SDSS RM, FITS with 276 col. Including vectors
9,5	3 cat. ASCII From author's archive + 9258 LC	13,5	10 tables in ASCII with bib. references, Abbreviated names+comp. not always the same between tables Not always coo.
7	VLA FF survey 14 tables ASCII (inc. from online archive) + Images FITS	11,3	SDSS-IV MaNGA FITS with 535 columns
6	3 tables MRTs + 2 behind Fig. +1 ASCII added for links toward spectra One name modified (exchange with author)	10,5	7 tables in ASCII Script to add coo in a table from another table of the same paper (not clear otherwise)
6	4 tables MRTs (T1 ~220 col.) + Table of 846 references	9	The XMM-SERVS survey 3 MRTs (inc. 206 & 197 col.)
6	4 tables (1 MRT, 3 ASCII) Curation of T2 complicated : wrong coo ; different values for NULL. 145 sources without RA (author has never answered)	8,3	IDEOS 1 table MRT , 232 col. SimbadName for 3559 sources
5,5	3 small Tables in MRTs Issues to link tables + SED Data presented in another paper + erratum (2 tables to add – via script; paste impossible)	8	22 tables en MRTs Names between tables for the same objects are not identical & not well written => misprints/comp. missing (diverse exchanges with the author)
5	1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII	7,5	BASS XXII. BASS DR2 AGN cat. 4 other papers of the same series to do. Flags added to link other tables let in the paper 2 MRTs tables + 1 ASCII Retrieving names from SWIFT.
5	Cosmicflows-4 2 MRTs (inc. 85 col). No coo. Retrieved from LEDA, 2MASX... => merge/lds added in SIMBAD + rewriting names for ~300 galaxies	7	BDKP. Paper V. 7 tables (3 MRTs, 4 ASCII) Names between tables => misprints – diverse exchanges with the author
5	1 MRT (~90 col.) No coo. 2 col. Simbad for 6393 pl. + stars	6	8 tables (6 en MRTs + tables added: notes & abund) SimbadName



Workflow insights – difficult cases

Number of columns per table increase => each column takes time:

- min/max values which should be coherent + only one NULL value available...
- Standard unit (if relevant)
- Only one explanation (no if there is ... then this is ... otherwise this is ... unless there is a flag...)
- See Point 2 of « Make your data visible »
- ucd1+
- Origin of the parameter (observation, ref.)
- Explanation of all codes/flags
- Origin of each IDs & SIMBAD nomenclature if relevant
- Which column(s) to display by default

What helps :

MRT format

Stick to one homogeneous type of data per column.

Stick to one NULL value per column

ApJ		ApJS	
Total hours	Catalog	Total hours	Catalog
24,5	RELICS, 41 catalogs ASCII (from MAST)	17	SDSS RM, FITS with 276 col. Including vectors
9,5	3 cat. ASCII From author's archive + 9258 LC	13,5	10 tables in ASCII with bib. references, Abbreviated names+comp. not always the same between tables Not always coo.
7	VLA FF survey 14 tables ASCII (inc. from online archive) + Images FITS	11,3	SDSS-IV MaNGA FITS with 535 columns
6	3 tables MRTs + 2 behind Fig. +1 ASCII added for links toward spectra One name modified (exchange with author)	10,5	7 tables in ASCII Script to add coo in a table from another table of the same paper (not clear otherwise)
6	4 tables MRTs (T1 ~220 col.) + Table of 846 references	9	The XMM-SERVS survey 3 MRTs (inc. 206 & 197 col.)
6	4 tables (1 MRT, 3 ASCII) Curation of T2 complicated : wrong coo ; different values for NULL. 145 sources without RA (author has never answered)	8,3	IDEOS 1 table MRT , 232 col. SimbadName for 3559 sources
5,5	3 small Tables in MRTs Issues to link tables + SED Data presented in another paper + erratum (2 tables to add – via script; paste impossible)	8	22 tables en MRTs Names between tables for the same objects are not identical & not well written => misprints/comp. missing (diverse exchanges with the author)
5	1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII	7,5	BASS XXII. BASS DR2 AGN cat. 4 other papers of the same series to do. Flags added to link other tables let in the paper 2 MRTs tables + 1 ASCII Retrieving names from SWIFT.
5	Cosmicflows-4 2 MRTs (inc. 85 col). No coo. Retrieved from LEDA, 2MASX... => merge/lds added in SIMBAD + rewriting names for ~300 galaxies	7	BDKP. Paper V. 7 tables (3 MRTs, 4 ASCII) Names between tables => misprints – diverse exchanges with the author
5	1 MRT (~90 col.) No coo. 2 col. Simbad for 6393 pl. + stars	6	8 tables (6 en MRTs + tables added: notes & abund) SimbadName



Workflow insights – difficult cases



No coordinates & link toward SIMBAD

=> Required for VizieR

Helps to detect misprints

- Find objects/coo in SIMBAD (potentially complete SIMBAD by adding IDs, data, merging... + complexity increased with number of rows)

- If new objects, find coo in other catalogs, other tables not in VizieR/not online...

What helps :

See point 1 of « Make your data visible »:
follow the IAU recommendation for non-altered/truncated names + coo + 2d name if available

ApJ		ApJS	
Total hours	Catalog	Total hours	Catalog
24,5	RELICS, 41 catalogs ASCII (from MAST)	17	SDSS RM, FITS with 276 col. Including vectors
9,5	3 cat. ASCII From author's archive + 9258 LC	13,5	10 tables in ASCII with bib. references, Abbreviated names+comp. not always the same between tables Not always coo.
7	VLA FF survey 14 tables ASCII (inc. from online archive) + Images FITS	11,3	SDSS-IV MaNGA FITS with 535 columns
6	3 tables MRTs + 2 behind Fig. +1 ASCII added for links toward spectra One name modified (exchange with author)	10,5	7 tables in ASCII Script to add coo in a table from another table of the same paper (not clear otherwise)
6	4 tables MRTs (T1 ~220 col.) + Table of 846 references	9	The XMM-SERVS survey 3 MRTs (inc. 206 & 197 col.)
6	4 tables (1 MRT, 3 ASCII) Curation of T2 complicated : wrong coo ; different values for NULL. 145 sources without RA (author has never answered)	8,3	IDEOS 1 table MRT , 232 col. SimbadName for 3559 sources
5,5	3 small Tables in MRTs Issues to link tables + SED Data presented in another paper + erratum (2 tables to add – via script; paste impossible)	8	22 tables en MRTs Names between tables for the same objects are not identical & not well written => misprints/comp. missing (diverse exchanges with the author)
5	1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII	7,5	BASS XXII. BASS DR2 AGN cat. 4 other papers of the same series to do. Flags added to link other tables let in the paper 2 MRTs tables + 1 ASCII Retrieving names from SWIFT.
5	Cosmicflows-4 2 MRTs (inc. 85 col). No coo. Retrieved from LEDA, 2MASX... => merge/lds added in SIMBAD + rewriting names for ~300 galaxies	7	BDKP. Paper V. 7 tables (3 MRTs, 4 ASCII) Names between tables => misprints – diverse exchanges with the author
5	1 MRT (~90 col.) No coo. 2 col. Simbad for 6393 pl. + stars	6	8 tables (6 en MRTs + tables added: notes & abund) SimbadName



Workflow insights – difficult cases

Corrections

=> Over 3 months in 2018, corrections in ApJ/ApJS leading to exchanges with authors were ~30 %

=> The main corrections concern identifiers, missing/erroneous coordinates, units, odd/redundant values...

=> This adds a further delay to the catalog release

ApJ		ApJS	
Total hours	Catalog	Total hours	Catalog
24,5	RELICS, 41 catalogs ASCII (from MAST)	17	SDSS RM, FITS with 276 col. Including vectors
9,5	3 cat. ASCII From author's archive + 9258 LC	13,5	10 tables in ASCII with bib. references, Abbreviated names+comp. not always the same between tables Not always coo.
7	VLA FF survey 14 tables ASCII (inc. from online archive) + Images FITS	11,3	SDSS-IV MaNGA FITS with 535 columns
6	3 tables MRTs + 2 behind Fig. +1 ASCII added for links toward spectra One name modified (exchange with author)	10,5	7 tables in ASCII Script to add coo in a table from another table of the same paper (not clear otherwise)
6	4 tables MRTs (T1 ~220 col.) + Table of 846 references	9	The XMM-SERVS survey 3 MRTs (inc. 206 & 197 col.)
6	4 tables (1 MRT, 3 ASCII) Curation of T2 complicated : wrong coo ; different values for NULL. 145 sources without RA (author has never answered)	8,3	IDEOS 1 table MRT , 232 col. SimbadName for 3559 sources
5,5	3 small Tables in MRTs Issues to link tables + SED Data presented in another paper + erratum (2 tables to add – via script; paste impossible)	8	22 tables en MRTs Names between tables for the same objects are not identical & not well written => misprints/comp. missing (diverse exchanges with the author)
5	1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII	7,5	BASS XXII. BASS DR2 AGN cat. 4 other papers of the same series to do. Flags added to link other tables let in the paper 2 MRTs tables + 1 ASCII Retrieving names from SWIFT.
5	Cosmicflows-4 2 MRTs (inc. 85 col). No coo. Retrieved from LEDA, 2MASX... => merge/lds added in SIMBAD + rewriting names for ~300 galaxies	7	BDKP. Paper V. 7 tables (3 MRTs, 4 ASCII) Names between tables => misprints – diverse exchanges with the author
5	1 MRT (~90 col.) No coo. 2 col. Simbad for 6393 pl. + stars	6	8 tables (6 en MRTs + tables added: notes & abund) SimbadName



Workflow insights – Value added

5,5

3 small Tables in MRTs
Issues to link tables + SED
Data presented in another paper +
erratum (2 tables to add – via script;
paste impossible)

- Show the target form
- Show constraint information

The 6 columns in **color** are computed by VizieR, and are **not part of the original data**.

[J/ApJ/879/131/psbgal](#) UV-FIR obs. of post-starburst galaxies & dust masses (Li+, 2019)

[Post annotation](#)

Post-starburst galaxies properties and archival UV-FIR photometry (Tables 1, 2, 4 and Tables 6+7 from Smercina et al. 2018ApJ...855...51S) (58 rows) [METAtab] [METAcola] [stats]

meta

[ReadMe+ftp](#)
2019ApJ...879..131L



spectrum/SED

[start AladinLite](#)

[plot the output](#)

[query using TAP/SQL](#)

J/ApJ/879/131/psbgal metadata

Metadata assigned by the CDS

Position used for search

System=FK5/2000
Ref. columns : DEJ2000, RAJ2000

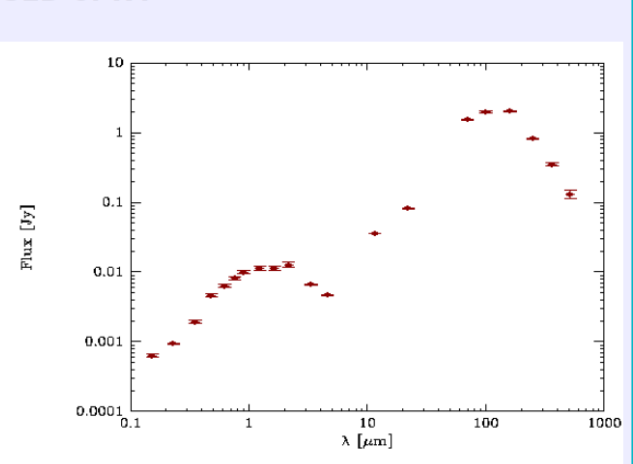
Photometry

column Fu, system=SDSS/u
 column Fg, system=SDSS/g
 column Fr, system=SDSS/r
 column Fi, system=SDSS/i
 column Fz, system=SDSS/z
 column FJ, system=2MASS/J
 column FH, system=2MASS/H
 column FKs, system=2MASS/Ks
 column F70, system=Herschel/PACS/70
 column F100, system=Herschel/PACS/100
 column F160, system=Herschel/PACS/160
 column F250, system=Herschel/SPIRE/250
 column F350, system=Herschel/SPIRE/350
 column F500, system=Herschel/SPIRE/500
 column F3.4, system=WISE/W1
 column F4.6, system=WISE/W2
 column F12, system=WISE/W3
 column F22, system=WISE/W4
 column FFUV, system=GALEX/FUV
 column FNUV, system=GALEX/NUV

<u>DEJ2000</u> "d:m:s"	<u>ID</u>	<u>RAJ2000</u> deg	<u>DEJ2000</u> deg	<u>FFUV</u> mJy	<u>F500</u> mJy	<u>SED</u>	<u>f_ID</u>	<u>SCat</u>	<u>Sloan</u>	<u>SimbadName</u>
+57 52 58.512	R1	233.13199	+57.88292	5.8e-01	60.90	SED				
+20 01 20.496	R2	228.95127	+20.02236	2.3e-01	105.00	SED		1	Sloan	SDSS J151548.33+200120.8
+16 43 46.848	R3	225.40127	+16.72968	6.2e-01		SED		1	Sloan	SDSS J150136.30+164346.9
+40 20 42.756	R4	246.45527	+40.34521	6.3e-01	131.00	SED		1	Sloan	SDSS J162549.26+402042.8
+14 03						SED		1	Sloan	SDSS J161735.41+140308.2
+41 40						SED		1	Sloan	SDSS J165141.69+414006.1
+13 51						SED		1	Sloan	SDSS J163758.87+135133.8
+52 29						SED		1	Sloan	SDSS J155816.43+522921.3
+22 23						SED		1	Sloan	SDSS J162842.96+222349.8
+17 20						SED	a	1	Sloan	[DYZ2015] EAH01
+18 40						SED	a	0	Sloan	[DYZ2015] EAH02
+17 33						SED	a	1	Sloan	[DYZ2015] EAH03
+00 32						SED	a	1	Sloan	[DYZ2015] EAH04
+39 04						SED	a	1	Sloan	[DYZ2015] EAH05
+31 22						SED	a	0	Sloan	[DYZ2015] EAH06
+11 33						SED	a	0	Sloan	[DYZ2015] EAH07
+02 30						SED	a	0	Sloan	[DYZ2015] EAH08
+37 33						SED	a	0	Sloan	[DYZ2015] EAH09
+21 07						SED	a	1	Sloan	[DYZ2015] EAH10
+05 59						SED	a	1	Sloan	[DYZ2015] EAH11
+13 16						SED	a	0	Sloan	[DYZ2015] EAH12
+22 09 47.448	EAH13	155.50328	+22.16318	2.0e-02		SED	a	1	Sloan	[DYZ2015] EAH13
+64 17 56.508	EAH14	178.27686	+64.29903	3.7e-03		SED	a	3	Sloan	[DYZ2015] EAH14
24 10 52 20.4480	+05 49 41.592	EAH15	163.08520	+05.82822	1.5e-02	SED	a	1	Sloan	[DYZ2015] EAH15
25 09 26 57.6888	+42 31 36.624	EAH16	141.74037	+42.52684	2.0e-02	SED	a	1	Sloan	[DYZ2015] EAH16
26 12 44 51.6936	-01 45 35.640	EAH17	191.21539	-01.75990	4.4e-03	SED	a	1	Sloan	[DYZ2015] EAH17

Display data from French+, 2018, J/ApJ/862/2 and Alatalo+, 2016, J/ApJS/224/38 within 0.5"

SED of R4



[Postscript Figure](#)

[Data as a Table](#)

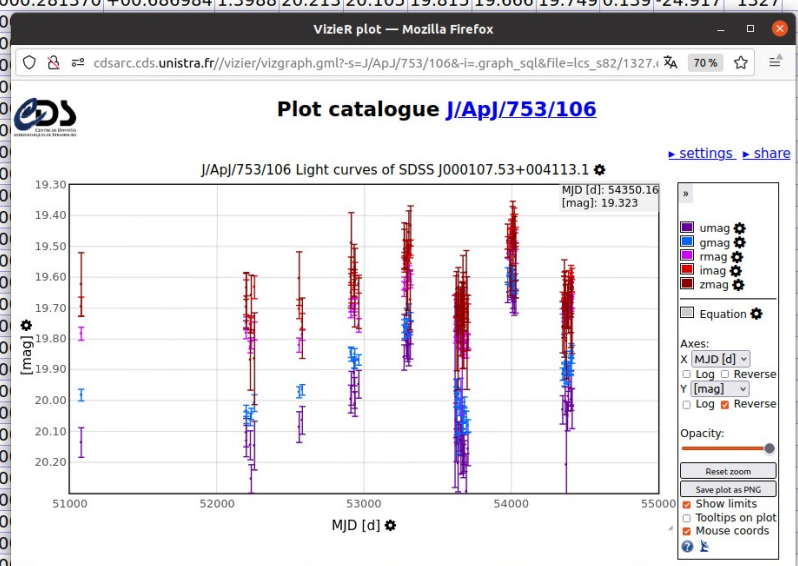


Workflow insights – Value added



5 1 MRT + 3 Data behind Figures inc. Individual Sp (26 MRTs) + 1 ASCII

RAJ2000 deg	DEJ2000 deg	z	umag	gmag	rmag	imag	zmag	Au mag	iMag mag	DBID	SDRSID	iMag mag	MBH [Msun]	logLbol [10-7W]	S11 LC	SimbadName
002.169302	+01.238649	1.0730	20.793	20.469	20.197	20.040	20.000	0.116	-23.901	70	301	-24.181			S11 LC	SDSS J000840.63+011419.2
001.091028	+00.962126	0.7867	20.790	20.183	19.849	19.818	19.430	0.183	-23.399	98					S11 LC	SDSS J00421.84+005743.6
000.331289	+00.177230	1.6199	20.892	20.554	20.431	20.199	20.099	0.154	-24.735	233					S11 LC	SDSS J000119.50+001037.9
001.364696	-00.098956	0.6125	20.098	19.722	19.784	19.485	19.541	0.178	-23.121	1018	190	-24.045	45.433		S11 LC	SDSS J000527.52-000556.2
000.221552	-00.292485	2.7563	20.707	19.663	19.610	19.705	19.529	0.174	-26.451	1310	36	-26.974	9.361	46.760	S11 LC	SDSS J000053.17-001732.9
000.281370	+00.686984	1.3988	20.213	20.105	19.815	19.666	19.749	0.139	-24.917	1327	49	-25.288	8.817	46.071	S11 LC	
											370	-27.856			S11 LC	
											229	-24.898	8.666	45.862	S11 LC	
											699	-25.695			S11 LC	
											578	-25.581			S11 LC	
											237	-27.684	9.591	46.989	S11 LC	
											13	-25.495	8.664	46.241	S11 LC	
											83	-25.560			S11 LC	
											564	-26.598	9.210	46.597	S11 LC	
											579	-25.739	8.885	46.291	S11 LC	
											556	-24.652	8.374	45.920	S11 LC	
											565	-24.361	8.415	45.833	S11 LC	
											574	-26.822			S11 LC	
											349	-25.560	8.549	46.250	S11 LC	
											555	-25.417	8.593	46.105	S11 LC	
											676	-23.600			S11 LC	
											680	-25.888			S11 LC	
											158	-25.326	9.189	46.107	S11 LC	
											621	-25.342	8.476	46.223	S11 LC	
											667	-25.431			S11 LC	
											519	-25.517	8.737	46.005	S11 LC	
											196	-23.487	8.049	45.327	S11 LC	
											407	-25.294	8.748	46.125	S11 LC	
											231	-26.337	8.997	46.592	S11 LC	
											530	-27.489	9.180	46.942	S11 LC	
											150	-25.563	8.371	46.194	S11 LC	
											652	-24.793	8.893	46.096	S11 LC	
											626	-26.408	9.099	46.549	S11 LC	
											183	-25.046	8.972	45.923	S11 LC	
											138	-25.515	8.140	46.117	S11 LC	



Full	Date "Y:M:D"	f	MJD d	delT d	Tel	Inst	u	Exp s	f	sp
1	2018-03-06		58183.5	72	MDM Hiltner 2.4m	OSMOS		600		
2	2018-03-08		58185.5	74	MDM Hiltner 2.4m	OSMOS		1800		
3	2018-03-09		58186.5	75	MDM Hiltner 2.4m	OSMOS		2700		
4	2018-03-13		58190.0	79	Liverpool Telescope 2m	SPRAT		600		
5	2018-03-23	*	58200.5	89	APO 3.5m	DIS		300		
6	2018-04-23		58231.6	119	Las Cumbres 2m	FLOYDS		1800		
7	2018-04-24		58232.5	120	Lick Shane 3m	Kast	~	1500	d	
8	2018-05-07		58245.6	133	Las Cumbres 2m	FLOYDS		1800		
9	2018-05-14		58252.5	140	Las Cumbres 2m	FLOYDS		1800		
10	2018-05-28		58266.5	153	Las Cumbres 2m	FLOYDS		1800		
11	2018-06-03		58272.6	159	Las Cumbres 2m	FLOYDS		1800		
12	2018-06-11		58280.6	167	Las Cumbres 2m	FLOYDS		1800		
13	2018-06-13	*	58282.5	169	APO 3.5m	DIS		600		
14	2018-06-24		58293.3	180	Las Cumbres 2m	FLOYDS		1800		
15	2018-06-28		58297.5	184	Las Cumbres 2m	FLOYDS		3600		
16	2018-07-06		58305.5	192	Las Cumbres 2m	FLOYDS		3600		

9,5 3 cat. ASCII From author's archive + 9258 LC



Workflow insights – Value added

7

VLA FF survey
14 tables ASCII (inc. from online archive)
+ Images FITS

J/ApJ/910/105 VLA Frontier Fields survey for 3 MACS clusters (Heywood+, 2021) [sggda/fits](#)

[Similar Catalogs](#) [2021ApJ...910..105H](#) [ReadMe+ftp](#)

- [J/ApJ/910/105/table1](#) (c)Coordinates and calibrators for each of the 3 target clusters, as well as the on-source integration times for each of the configuration/band pairings (3 rows)
- [J/ApJ/910/105/extend](#) (c)Positions and integrated flux densities of the extended radio sources in MACS J0416.1-2403 (Table 4), MACS J0717.5+3745 (Table 5) and MACS J1149.5+2223 (Table 6) (66 rows)
- [J/ApJ/910/105/opt](#) (c)3GHz-detected compact radio sources with optical counterparts in MACSJ0416.1-2403, MACSJ0717.5+3745 and MACSJ1149.5+2223 (1296 rows)
- [J/ApJ/910/105/rad](#) (c)3GHz-detected compact radio sources without optical counterpart in MACSJ0416.1-2403, MACSJ0717.5+3745 and MACSJ1149.5+2223 (670 rows)
- [J/ApJ/910/105/cband](#) (c)C-band (6GHz) compact component morphology for sources in MACSJ0416.1-2403, MACSJ0717.5+3745 and MACSJ1149.5+2223 (262 rows)
- [J/ApJ/910/105/table3](#) (c)Magnifications and demagnified integrated flux densities, peak brightnesses, and effective noise levels for the 13 lensed compact radio sources presented in Figure 6 (13 rows)



ing the VizieR catalogues

series, SED) which comes from publications. This tool is supervised by the CDS documentalists team (see the



A, SSA, ObsTAP) and can so be queried by VO
ObsCore has been chosen for the documentation.

Simple search [ObsTAP Query](#)

Q Search by position : 109.3922 +37.7473 radius 1 deg

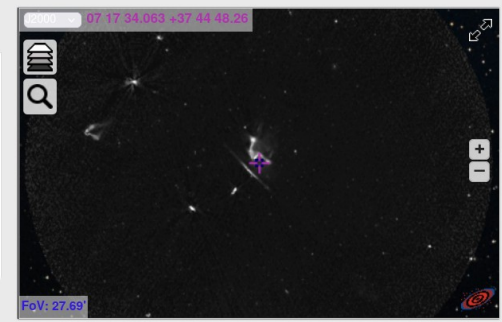
Q Search by spectral band : min max nm

Q Search by time data : start stop (MJD)

Q Search by catalog: J/ApJ/910/105 Q Identifier:

Spectrum / Time series Image

500 entries max



Show 10 entries

4 entries

Preview	Target	Data collection	Ra	Dec	Band min (nm)	Band max (nm)	Begin time (MJD)	End time (MJD)	Facility			
	MACSJ0717.5+37	J/ApJ/910/105	109.392	37.747	99,964,154.718	99,964,154.718	56,601.217		EVLA			Header
	MACSJ0717.5+37	J/ApJ/910/105	109.392	37.747	99,964,154.718	99,964,154.718	56,601.217		EVLA			Header
	MACSJ0717.5+37	J/ApJ/910/105	109.392	37.747	49,973,745.624	49,973,745.624	56,746.880		EVLA			Header
	MACSJ0717.5+37	J/ApJ/910/105	109.392	37.747	49,973,745.624	49,973,745.624	56,746.880		EVLA			Header



In a nutshell...



- **Balance between quantity and quality** to keep (not always easy...)
- **Half of the annual ingestion** in VizieR is for **AAS journals** (for 1-2 documentalists), in VizieR, **currently 41% of the catalogs** (including large catalogs as Gaia, SDSS...) are AAS journals (it should increase if the ingestion rate stays the same).
- **Time spent on one catalog is really variable**
 - It depends of the **content of the catalog** inherent to the data :
Many tables, many number of columns, many rows, many IDs, many filters, many associated data... will inevitably increase the processing time
 - It depends of the **quality of the catalog** :
=> MRT files are time-savers (especially combined with getapj)
=> Following the Best practice rules (Chen et al. 2022) or at least the 3 points given in « Make your data visible » flyer
- There will always be an **incompressible processing time** inherent in the **value added** to catalogs:
 - Detailed explanations for each column (provenance/content), UCD, METAtag
 - Links between tables of the same catalog, with other catalogs, SIMBAD or other databases
 - Plots & specific treatment for associated data
 - Corrections
 - ...