



The Gaia Archive

A. Mora, J. Gonzalez-Núñez, J. Salgado,
R. Gutiérrez-Sánchez, J.C. Segovia, J. Duran
ESA-ESAC Gaia SOC and ESDC

IAU Symposium 330. Nice, France

ESA UNCLASSIFIED - For Official Use



European Space Agency

Outline



1. Introduction
2. The Gaia Archive
3. Tables: catalogue, cross-match and light curves
4. Archive demo
5. How to select a stellar cluster (basic query)
6. How to create an HR diagram (2MASS cross-match)
7. How to reconstruct an RR Lyrae light curve
8. Additional resources
9. Conclusions

1. Introduction

ESA Gaia Archive: key points



Code to the data (workflows)

Gaia Archive **contents**. Provided by DPAC

Catalogues (tables)

Pre-computed **cross-match** (pivot tables)

Light curves (one entry per point)

Gaia Archive **functionality**

ADQL: SQL dialect for queries → **reproducibility** → papers

TAP+ VO standard: web data **query**, **user tables**, **sharing**, ...

SAMP VO standard: interoperability → **3rd party** software

DR2+ evolution: ~PB size, parameters, spectra, epoch data, ...

Learning curve: reproducing DR1 Brown+ 2016 plots

Code to the data (workflows)



There is no correct workflow! if it works, it is good

Traditional (bring the data to the code)

Download tgas_source (~1GB), gaia_source (~1TB) and external catalogues

Ingest in local supercomputer. Do cross-match

Do science!

Gaia archive (towards bringing the **code to the data**)

Selection (ADQL query) at Gaia Archive → **Reproducibility!**

Refinement (in Archive): Xmatch, computations, external catalogues, share

Download and do science! → Data (size) reduction

2. The Gaia Archive

Gaia Archive contents: DR1



Where?

Main Gaia archive: ESAC

Partner data centres: AIP, ARI, ASI, BCN, CDS

Affiliated and other data centres, enthusiasts: tens

What?

tgas_source (~2M 5 parameter astrometry): **all**

gala_source (~1.1G positions): **most**

Variability (599 cepheids, 2595 RR Lyrae), QSOs (2191): **ESAC+**

Xmatch (2MASS, PPMXL, SDSS9 , UCAC4, URAT1, WISE): **ESAC**

External catalogues (for Xmatch): **ESAC+**

What else @ESAC?: TAP+ (data base, user space, share), cross-match

Total DR1: 1.6 TB

Gaia

•Gaia DR1 catalogue	1.1	billion rows
•TGAS	0.002	billion rows

External Catalogues

•Hipparcos & Hipparcos new red.	0.0012	billion rows
•IGSL (Initial Gaia Source List)	1.2	billion rows
•2MASS	0.47	billion rows
•Tycho2	0.0025	billion rows
•UCAC4	0.11	billion rows
•Hubble Source Catalogue v1.0	0.029	billion rows

Crossmatches

•Crossmatch tables between Hipparcos, 2MASS, Tycho2... and Gaia expressed as neighbourhood and best neighbour, e.g:		
•AllWise-Gaia neighbourhood	0.31	billion rows

11.59 billion rows

The User Interface



The screenshot shows the Gaia Archive web interface. At the top, there's a navigation bar with 'HOME', 'SEARCH', 'STATISTICS', 'VISUALIZATION', 'HELP', 'DOCUMENTATION', 'VOSPACE', 'SHARE', and 'ADMIN'. Below this is a 'Simple Form' tab and an 'ADQL Form' tab. The main content area includes a 'Job name' input field, a 'Query examples' section with a sample SQL query, and 'Reset Form' and 'Submit Query' buttons. A table of job results is displayed below, with columns for Job, Creation date, Num. rows, and Size. The table contains 10 rows of data. At the bottom, there are controls for 'Apply jobs filter', 'Select all jobs', and 'Delete selected jobs'. The footer includes copyright information and the version number (v1.1.0).

gaia archive

HOME SEARCH STATISTICS VISUALIZATION HELP DOCUMENTATION VOSPACE SHARE ADMIN

Simple Form **ADQL Form** Query Results

Job name:

Query examples

```
1 SELECT DISTANCE(POINT('ICRS',ra,dec), POINT('ICRS',266.41683,-29.00781)) AS dist, *
2 FROM gaiadr1.gaia_source
3 WHERE 1=CONTAINS(POINT('ICRS',ra,dec),CIRCLE('ICRS',266.41683,-29.00781, 0.08333333)) ORDER BY dist ASC
```

Reset Form Submit Query

Job	Creation date	Num. rows	Size
1478302193217O	05-Nov-2016, 00:29:53	1316	258 KB
xmatch_hipparcos_hubble_sc	28-Oct-2016, 12:26:29		0 KB
xmatch_hipparcos_hubble_sc	28-Oct-2016, 12:26:01		0 KB
1477493920727O	26-Oct-2016, 16:58:40	81445	14 MB
1477493900212O	26-Oct-2016, 16:58:20	500	93 KB
1477493890497O	26-Oct-2016, 16:58:10	500	92 KB
1477493063430O	26-Oct-2016, 16:44:23	647	186 KB
1477493049348O	26-Oct-2016, 16:44:09		0 KB
1477492042282O	26-Oct-2016, 16:27:22	488	142 KB

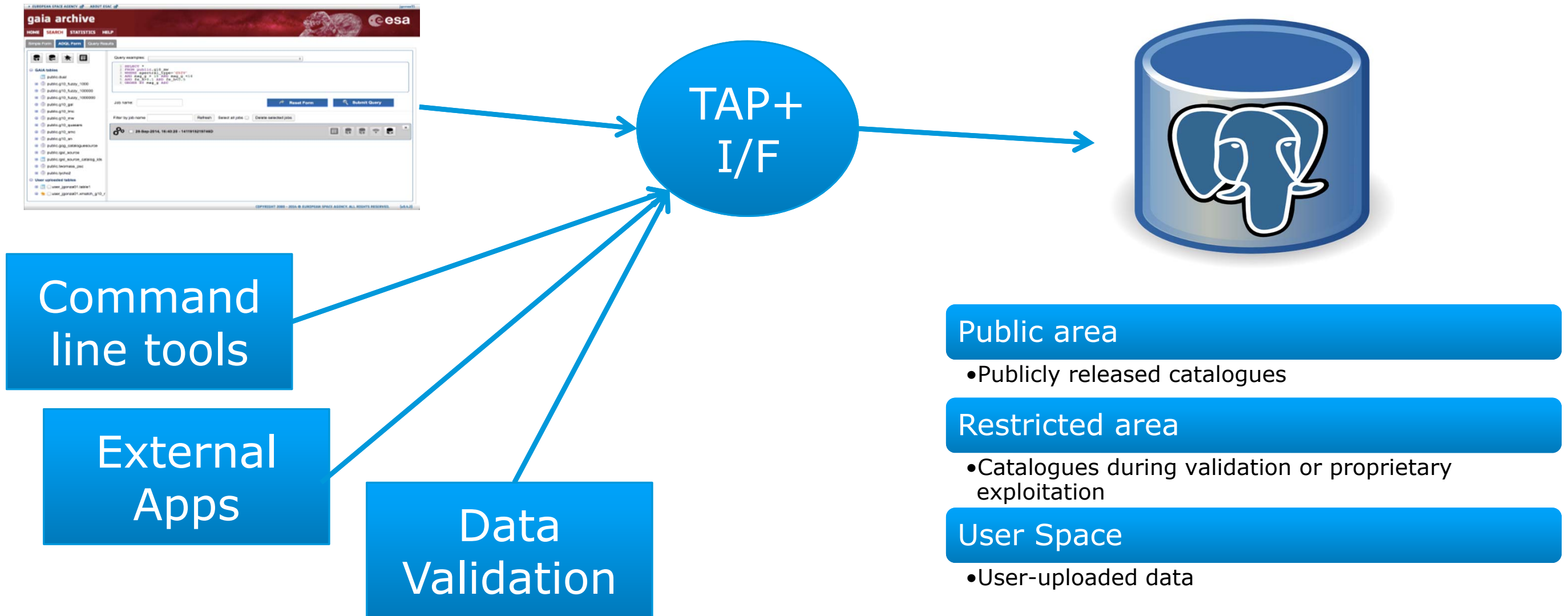
1-20 of 106

Apply jobs filter Select all jobs Delete selected jobs

COPYRIGHT 2000 - 2016 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED. (v1.1.0)

<http://archives.esac.esa.int/gaia>

Open APIs



Gaia Archive: TAP+ Schema



GAIA tables

- User tables
 - user_jsalgado.cata
 - user_jsalgado.dwarfs
 - user_jsalgado.early_stars
 - user_jsalgado.radial_velocities
 - user_jsalgado.t1404481668974d
 - user_jsalgado.visual_binaries
 - user_jsalgado.xmatch_cata_igsl_source
 - user_jsalgado.xmatch_igsl_source_dwarfs
 - user_jsalgado.xmatch_igsl_source_radial_velocities
 - user_jsalgado.xmatch_igsl_source_visual_binaries_2
- Shared to me (from satgaia)
 - user_satgaia.igsl_tmass_15s_bestneighbour
 - user_satgaia.igsl_tmass_15s_neighbourhood
 - user_satgaia.igsl_tmass_1s_bestneighbour
 - user_satgaia.igsl_tmass_1s_neighbourhood
 - user_satgaia.igsl_tmass_5s_bestneighbour
 - user_satgaia.igsl_tmass_5s_neighbourhood
 - user_satgaia.igsl_tycho2_15s_bestneighbour
 - user_satgaia.igsl_tycho2_15s_neighbourhood
 - user_satgaia.igsl_tycho2_1s_bestneighbour
 - user_satgaia.igsl_tycho2_1s_neighbourhood
 - user_satgaia.igsl_tycho2_5s_bestneighbour
 - user_satgaia.igsl_tycho2_5s_neighbourhood
 - user_satgaia.mdb00

GAIA Catalogue Upload

Select a file my_sources.vot

(*) Table name

Table description

Ra column name

Dec column name

(*) mandatory field

Persistent Upload

GAIA Cross-Match

Table A: user_jsalgado.cata Table B: public.igsl_source

Output table name: xmatch_cata_igsl_source

Radius: 1.0 (in arcseconds)

Server Crossmatch

GAIA Share Item

user_jsalgado.radial_velocities

Description:

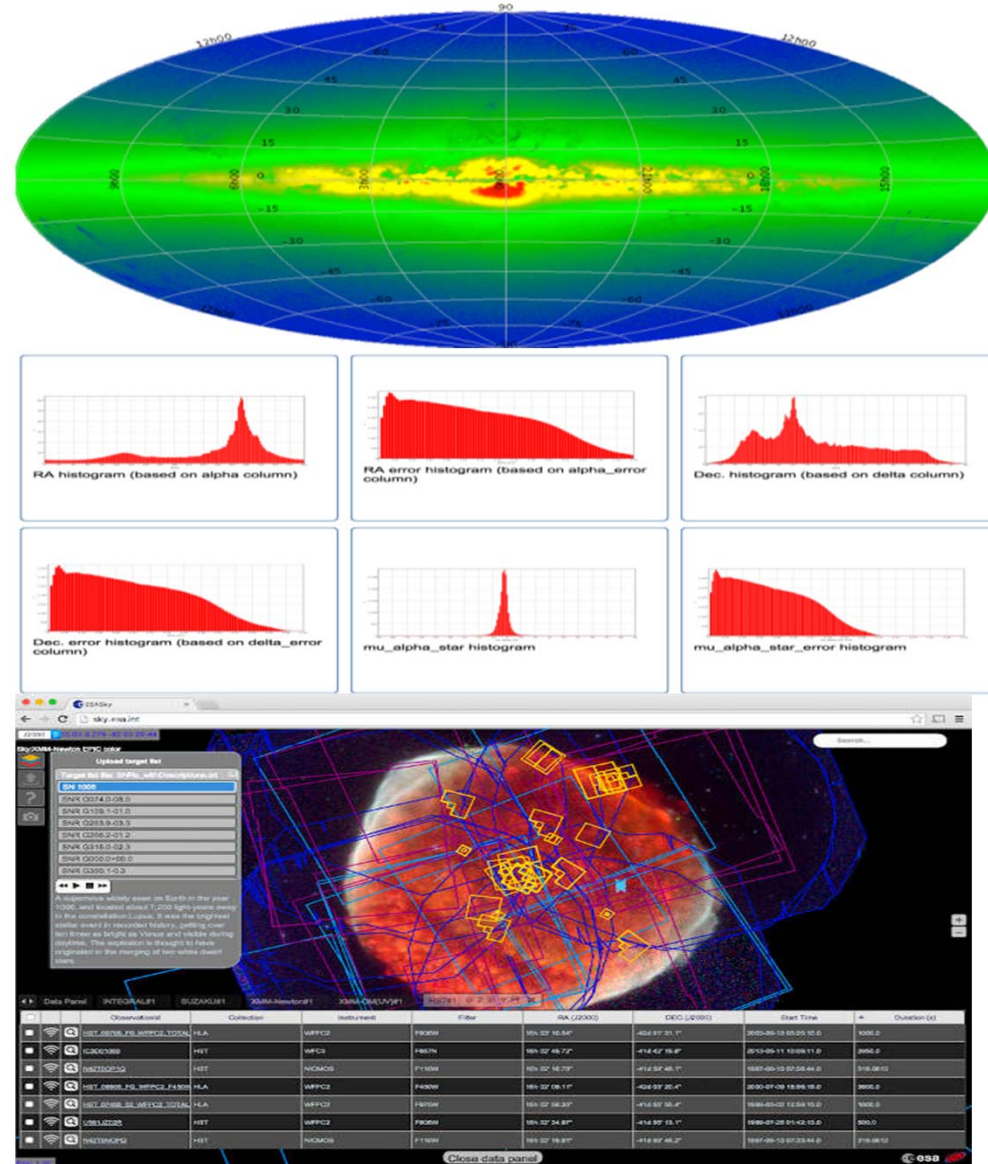
Shared to group: (None)

Share item to group: XMatch Group

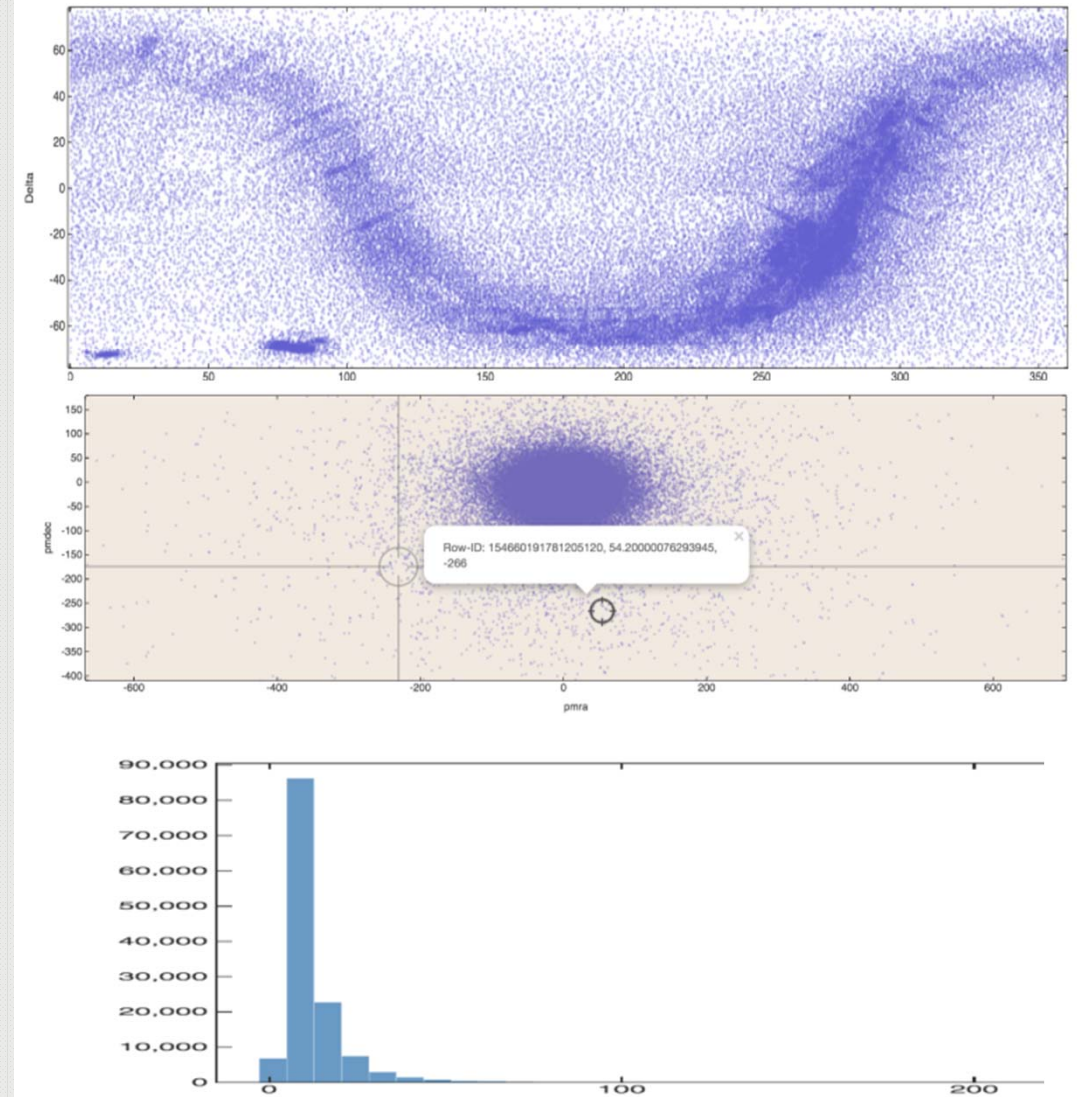
Table sharing



Visualization



- ❑ Visualization is a need
- ❑ Statistics provide holistic views. Big data techniques
- ❑ Validated static stats
- ❑ On-the-fly visualization (Lisbon University)
- ❑ ESASky integration
- ❑ Virtual Reality prototype



3. Tables: catalogue, cross-match and light curves

Gaia source: main table



EUROPEAN SPACE AGENCY ABOUT ESAC Alcione Mora (amora)

gaia archive

HOME **SEARCH** STATISTICS VISUALIZATION HELP DOCUMENTATION VOSPACE SHARE

Simple Form ADQL Form **Query Results**

No job id

Columns

solution_id	source_id	random_index	ref_epoch	ra	ra_error	dec	dec
			Time[Julian Years]	Angle[deg]	Angle[mas]	Angle[deg]	Angle[deg]
1635378410781933568	4211025012066010880	131410931	2015	290.2477074583974	93407	-6.331946167262529	2.53
1635378410781933568	4210922757482661248	856561327	2015	289.9415787981672	8632	-6.530635341768221	2.32
1635378410781933568	4210681380320174592	550360955	2015	294.267178592206	46088	-3.528292017227317	1.73
1635378410781933568	4211002098420565888	960015464	2015	289.5976320267487	16.94528450659924	-6.172833023239273	15.4
1635378410781933568	4210638499362703872	1004419411	2015	294.7500861075622	1.773955896855794	-3.636699368670485	1.58
1635378410781933568	4210847269136626304	557493500	2015	295.4382050078335	5.478051207928527	-2.6135820036757322	3.81
1635378410781933568						8415569492926	1.95
1635378410781933568						0625664414796	0.20
1635378410781933568						6804587432753	8.28
1635378410781933568						95552472134493	3.27
1635378410781933568	4210974851146901888	328770579	2015	289.70073316760386	2.958057395439199	-6.431398708912356	3.49
1635378410781933568	4210815658174807040	606797458	2015	294.9216310943188	15.266550937744302	-3.003298897829963	11.3
1635378410781933568	4210579125738498304	809666698	2015	296.07094455894094		629158240998	11.6
1635378410781933568	4210980653642584704	62717479	2015	289.9065131739514		728845502425	3.79
1635378410781933568	4210998864301014784	65705465	2015	289.67240825545065		576208601118	0.27

Units

Data: one entry (row) per object

Data model

Gaia Data Model Show query in ADQL form

1-20 of 20

Variables: light curve

One entry per object and epoch: 3194 stars, 233,181 rows
~Trillion size by DR5!

observation_time	g_flux	g_flux_error	g_magnitude
Time[Barycentric JD in TCB - 2455197.5 (day)]	Flux[e-/s]	Flux[e-/s]	Magnitude[mag]
1666.4708409132627	323.3406240378528	6.139588887158467	19.25061938345054
1666.646999930693	286.7358824320857	56.91077458377935	19.381064952793295
1666.7210076079693	255.0714586668238	4.621548747960895	19.508115399450226
1666.8971667883682	473.9340937184366	3.8638062521973255	18.835475183940574
1666.9711745074405	368.06374709170376	4.482043542948077	19.1099624559301
1667.147333701442	270.0869075966821	2.90891365826153	19.446011232919545
1667.2213414646678	262.13781543831874	3.894873668425852	19.478445874202386

Table: phot_variable_time_series_gfov

Pre-computed cross-matches



DR1 one data set. **Other catalogues:** astrometry, velocities, photometry, parameters

Ambiguity: resolution, wavelength, epoch, **proper motion**, ...

Many to many relationship → pivot tables needed → **traceability!**

Gaia Archive: **catalogue copies** and **pre-computed cross-matches**

Speed: indices. Efficiency: avoids ~100 M rows traffic

2MASS PSC: Skrutskie et al. 2006. 470,992,970 entries

AllWISE: Cutri et al. 2013, 747,634,026 entries

GSC2.3: Lasker et al. 2008. 945,592,683 entries

PPMXL: Roeser et al. 2010. 910,468,688 entries

SDSS DR9: Ahn et al. 2012. 469,029,929 entries

UCAC4: Zacharias et al. 2013. 113,728,883 entries

URAT-1: Zacharias et al. 2015. 228,276,482 entries

Xmatch: neighbourhood

gsc23_oid

source_id

original_ext_source_id

~Many to many relationship → double entry → one row per pair

185522891
185522891
185522891
185522891

6083712739771159808
6083712735454214528
6083712739771159424
6083712739771173760

S99R046506
S99R046506
S99R046506
S99R046506

186638244
186638244
186638244
186638244

6083712739771153360
6083717786336870400
6083717790654613888
6083717790654613760

ω Cen core

S99R121241
S99R121241
S99R121241
S99R121241

Xmatch: neighbourhood

gsc23_oid

source_id

original_ext_source_id

~Many to many relationship → double entry → one row per pair

186167520

6083518091833868416

S99R056045

186168205

6083518091833868416

S99R055731

186172303

6083518091833868416

S99R120298

186176636

6083518091833868416

S99R056043

ω Cen core

186208246

6083714492121489280

S99R055944

186211491

6083714492121489280

S99R139072

186211811

6083714492121489280

S99R139057

Xmatch: best neighbour

>0 → other Gaia sources share best neighbour

number_of_neighbours

number_of_mates

best_neighbour_multiplicity

8		1	1
8		0	1
8		0	1
8	ω Cen core	2	1
8		0	1
7		0	1
7		0	1

VO standards → 3rd party



Topcat (M. Taylor): Data query (TAP), reception (SAMP) & plotting

Table Access Protocol (TAP) Query

Window TAP Registry Edit Interop Help

Select Service Use Service Resume Job Running Jobs

Metadata

Find:

Name Descrip Or

gaiadr1 (8)

- gaiadr1.aux_qso_icrf2_r
- gaiadr1.cephheid
- gaiadr1.gaia_source
- gaiadr1.phot_variable_t
- gaiadr1.phot_variable_t
- gaiadr1.rriyrae
- gaiadr1.tgas_source
- gaiadr1.variable_summa

Service	Schema	Table	Columns	FKeys	Hints
		source_id	BIGINT	<input checked="" type="checkbox"/>	Unique source id
		ra	DOUBLE	<input checked="" type="checkbox"/>	Right ascension
		dec	DOUBLE	<input checked="" type="checkbox"/>	Declination
		l	DOUBLE	<input checked="" type="checkbox"/>	Galactic longitud
		b	DOUBLE	<input checked="" type="checkbox"/>	Galactic latitude
		ecl_lon	DOUBLE	<input checked="" type="checkbox"/>	Ecliptic longitud
		ecl_lat	DOUBLE	<input checked="" type="checkbox"/>	Ecliptic latitude
		parallax	DOUBLE	<input checked="" type="checkbox"/>	Parallax
		pmra	DOUBLE	<input checked="" type="checkbox"/>	Proper motion in
		pmdec	DOUBLE	<input checked="" type="checkbox"/>	Proper motion in

Service Capabilities

Query Language: ADQL-2.0 Max Rows: 100000 (default) Uploads: 1000krow/

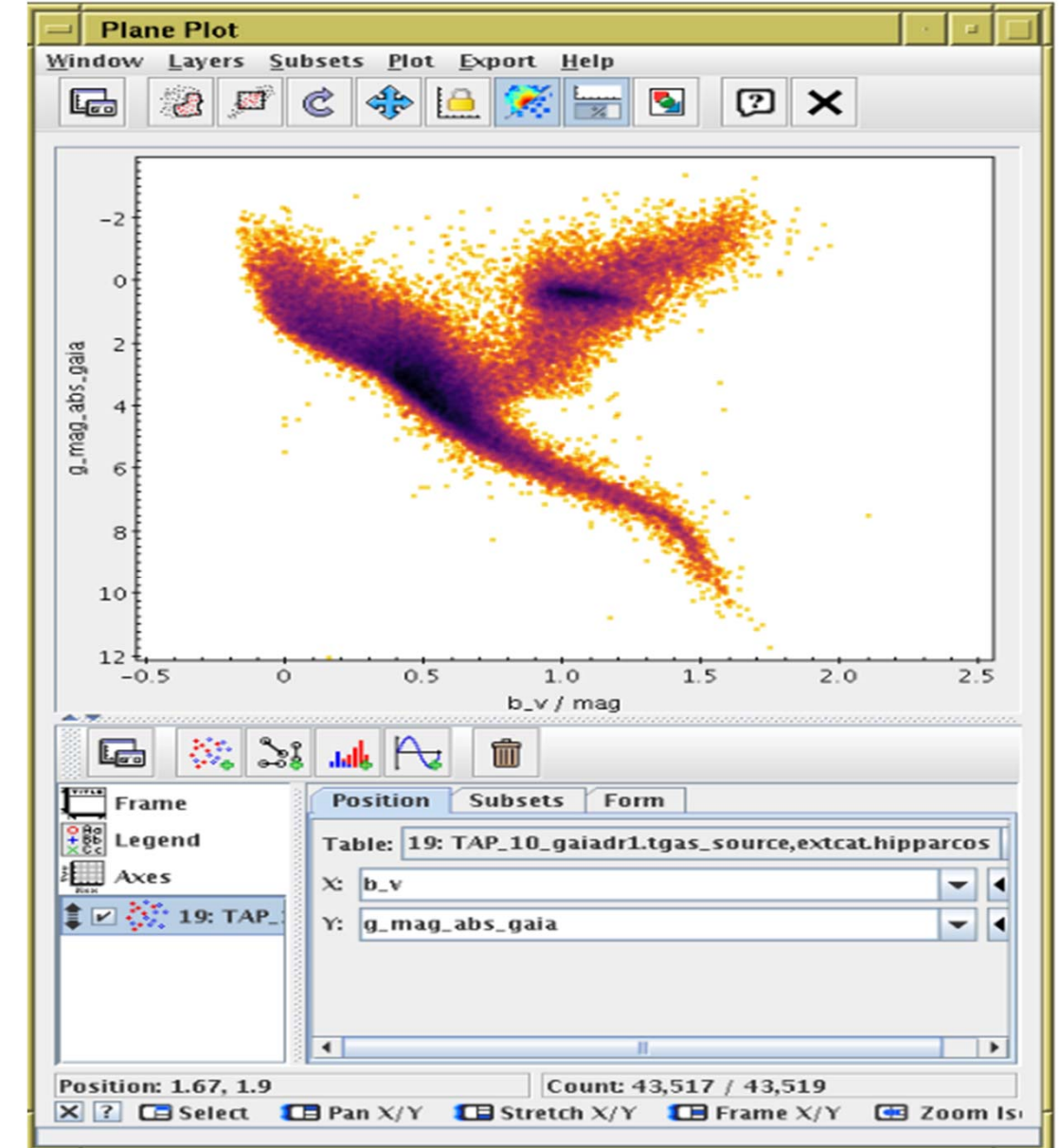
ADQL Text

Mode: Synchronous

```
SELECT TOP 50000
gaia.source_id,
gaia.hip,
gaia.phot,
gaia.phot,
hip.b_v
FROM gaiadr1.tgas
INNER JOIN e...
```

Get the 5 astrometric parameters
Cone Search with Galactic coordinates
Getting Healpix indices
3-D plot
Gaia DR1 - Color and magnitude 1/2
Gaia DR1 - Color and magnitude 2/2
Gaia DR1 - Density by magnitude
Gaia DR1 - Pleiades density by parallax

Run Query



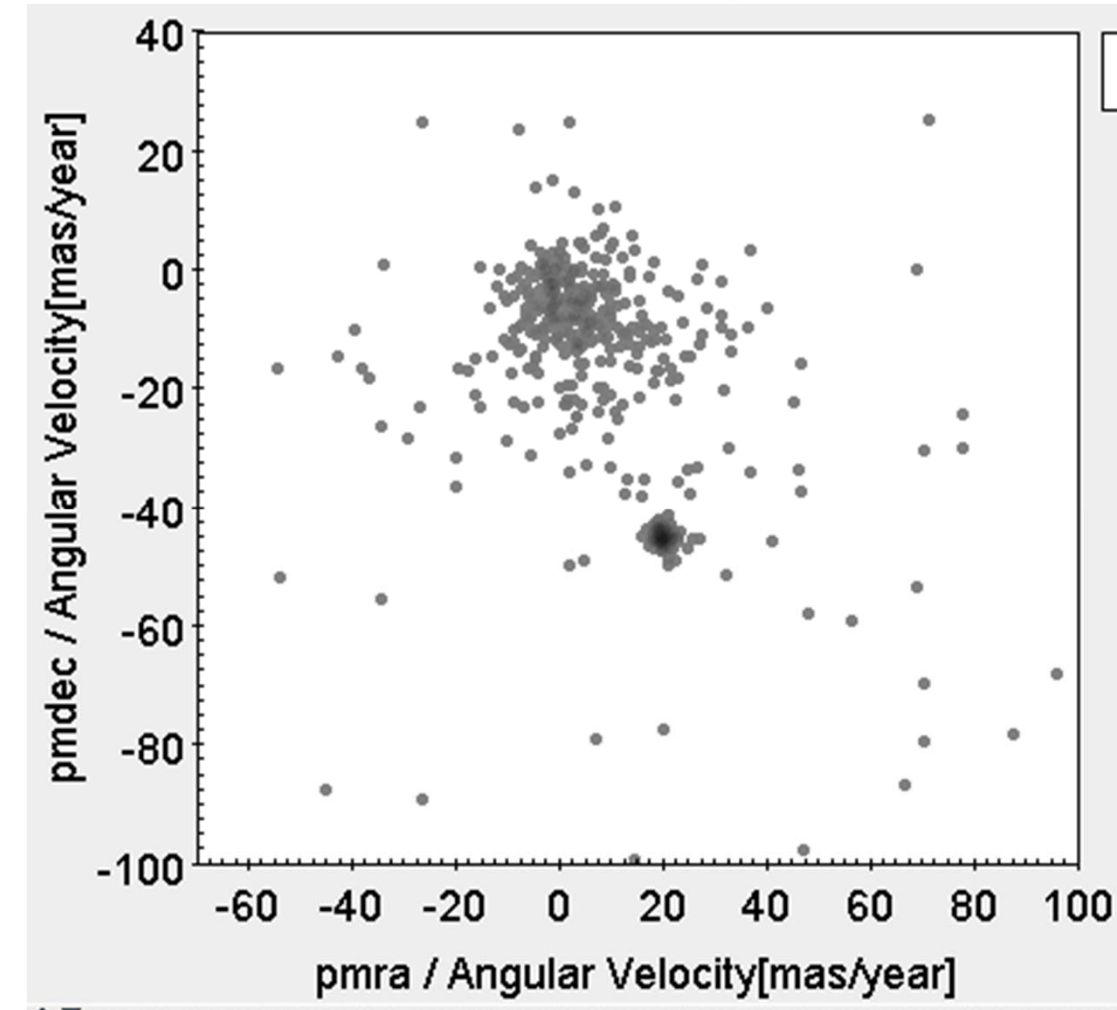
4. Archive demo

5. How to select a stellar cluster (basic query)

Tutorial: Pleiades gross selection



```
SELECT *  
FROM gaiadr1.tgas_source  
WHERE CONTAINS(  
    POINT('ICRS', ra, dec),  
    CIRCLE('ICRS', 56.75, 24.1167, 2)  
)=1
```

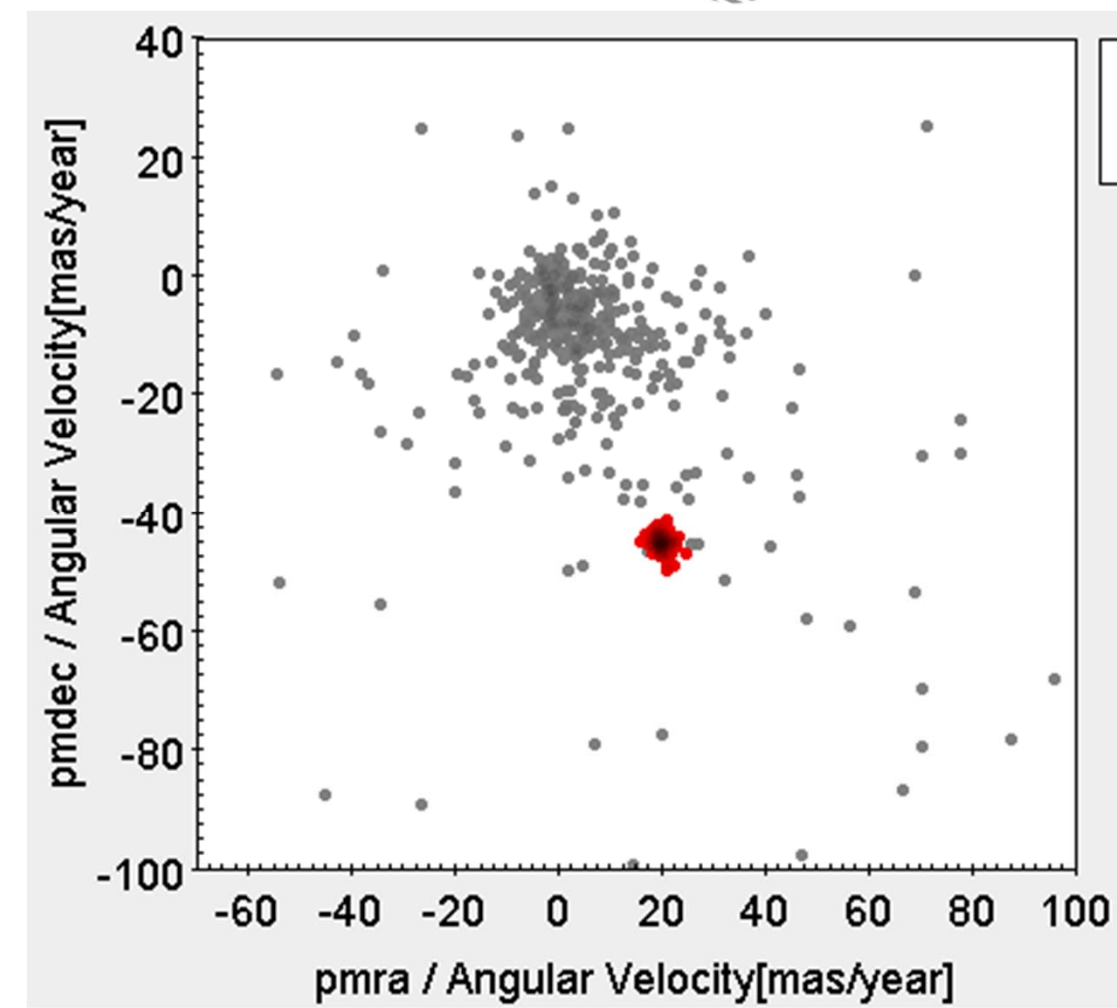


Cone search: 2° radius → Apparent in proper motion

Tutorial: Pleiades gross selection



```
SELECT *
FROM gaiadr1.tgas_source
WHERE CONTAINS(
    POINT('ICRS', ra, dec),
    CIRCLE('ICRS', 56.75, 24.1167, 2)
)=1
AND pmra IS NOT NULL AND pmra != 0
AND pmdec IS NOT NULL AND pmdec != 0
AND abs(pmra_error/pmra) < 0.10 AND abs(pmdec_error/pmdec) < 0.10
AND pmra BETWEEN 15 AND 25 AND pmdec BETWEEN -55 AND -40
```



PM filter

6. How to create an HR diagram (2MASS cross-match)

Brown+ 2016: HR diagram

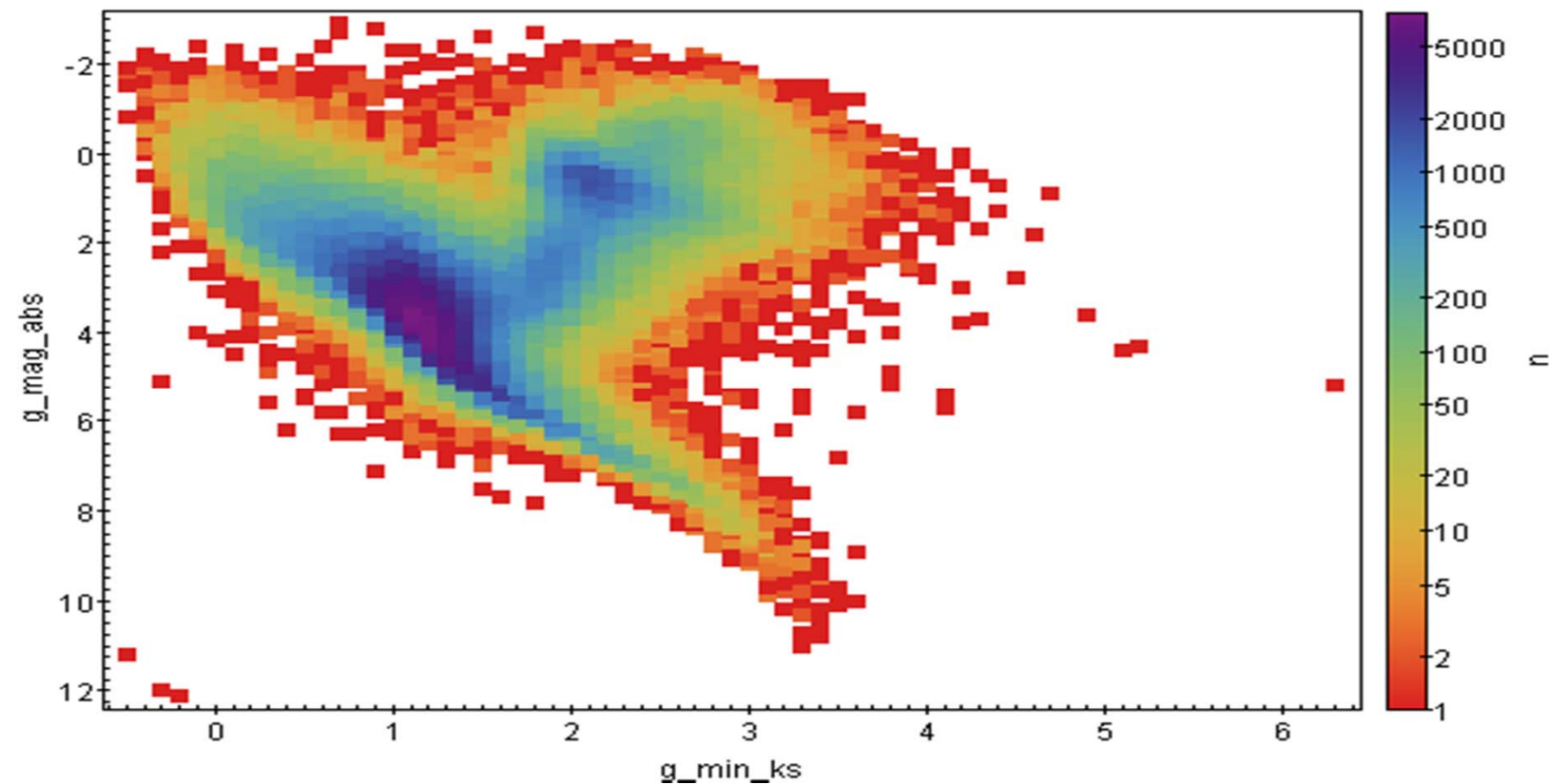
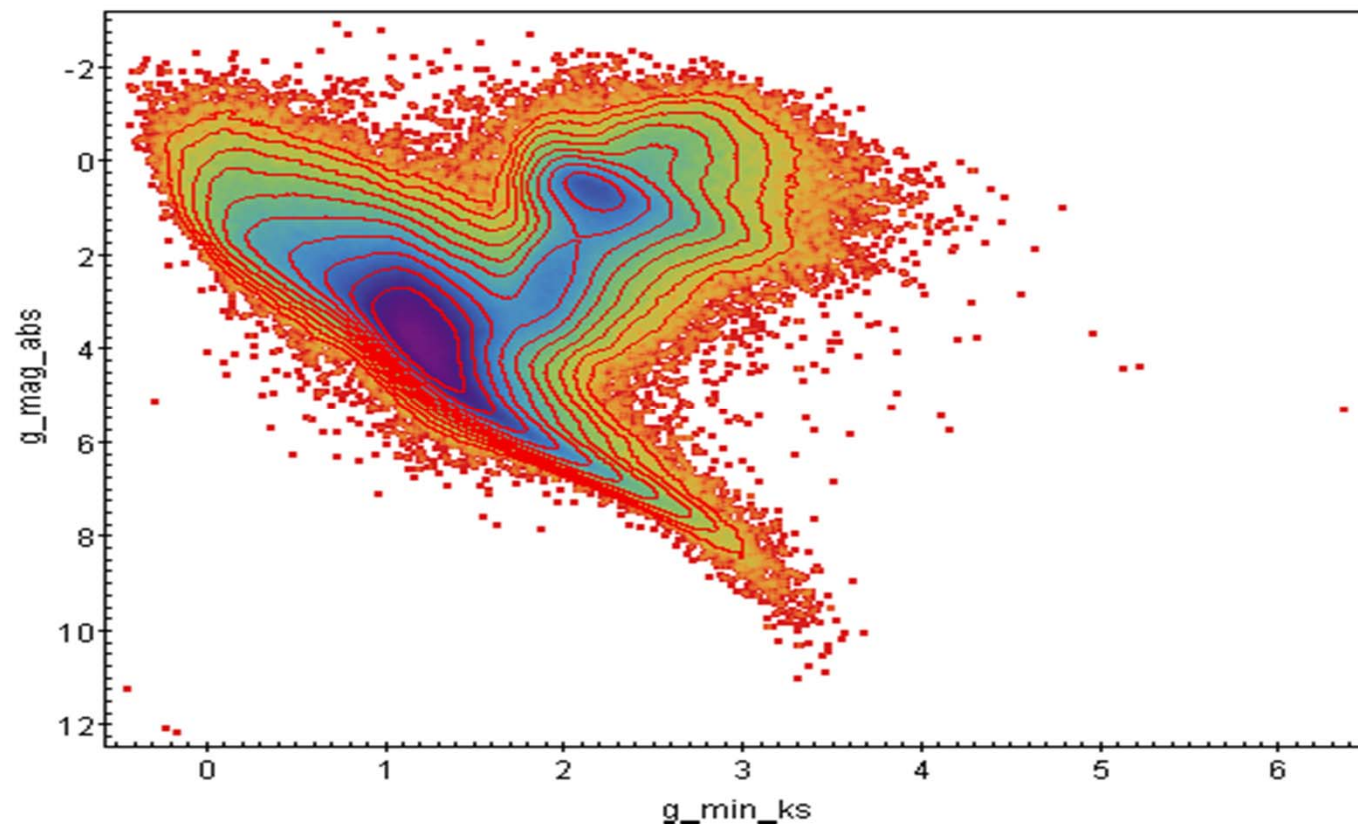


Plots are **reproducible!** ADQL queries in the paper

Uses Gaia parallaxes and fluxes → absolute magnitudes

Uses pre-computed cross-match and external catalogue: 2MASS

Histograms can be generated in the archive → Code to the data → DR2+



Brown+ 2016: HR diagram



```
select gaia.source_id,  
       gaia.phot_g_mean_mag + 5 * log10(gaia.parallax)- 10 as g_mag_abs,  
       gaia.phot_g_mean_mag-tmass.ks_m as g_min_ks  
from gaiadr1.tgas_source as gaia  
inner join gaiadr1.tmass_best_neighbour as xmatch  
       on gaia.source_id = xmatch.source_id  
inner join gaiadr1.tmass_original_valid as tmass  
       on tmass.tmass_oid = xmatch.tmass_oid  
where gaia.parallax/gaia.parallax_error >= 5 and  
       ph_qual = 'AAA' and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error  
               /gaia.phot_g_mean_flux,2) ) <= 0.05 and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error  
               /gaia.phot_g_mean_flux,2)  
       + power(tmass.ks_msigcom,2)) <= 0.05
```

**ADQL:
Learning curve**

Brown+ 2016: HR diagram



```
select gaia.source_id,  
       gaia.phot_g_mean_mag + 5 * log10(gaia.parallax)- 10 as g_mag_abs ,  
       gaia.phot_g_mean_mag - tmass.ks_m as g_min_ks  
from gaiadr1 tgas_source as gaia  
inner join gaiadr1 tmass_best_neighbour as xmatch  
       on gaia.source_id = xmatch.source_id  
inner join gaiadr1 tmass_original_valid as tmass  
       on tmass.tmass_oid = xmatch.tmass_oid  
where gaia.parallax/gaia.parallax_error >= 5 and  
       ph_qual = 'AAA' and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error  
               /gaia.phot_g_mean_flux,2) ) <= 0.05 and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error  
               /gaia.phot_g_mean_flux,2)  
       + power(tmass.ks_msigcom,2)) <= 0.05
```

ADQL: on the fly computation

Gaia + 2MASS:
Pre-computed Xmatch

ADQL: filters

Brown+ 2016: HR diagram



select

```
g_min_ks_index / 10 as g_min_ks,  
g_mag_abs_index / 10 as g_mag_abs,  
count(*) as n
```

from (

```
select gaia.source_id,  
       floor((gaia.phot_g_mean_mag+5*log10(gaia.parallax)-10) * 10) as g_mag_abs_index,  
       floor((gaia.phot_g_mean_mag-tmass.ks_m) * 10) as g_min_ks_index  
from gaiadr1.tgas_source as gaia  
inner join gaiadr1.tmass_best_neighbour as xmatch  
       on gaia.source_id = xmatch.source_id  
inner join gaiadr1.tmass_original_valid as tmass  
       on tmass.tmass_oid = xmatch.tmass_oid  
where gaia.parallax/gaia.parallax_error >= 5 and  
       ph_qual = 'AAA' and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)) <= 0.05 and  
       sqrt(power(2.5/log(10)*gaia.phot_g_mean_flux_error/gaia.phot_g_mean_flux,2)  
           + power(tmass.ks_msigcom,2)) <= 0.05
```

```
)as subquery
```

```
group by g_min_ks_index, g_mag_abs_index
```

**2D histogram in the archive:
no billion object download → DR2**

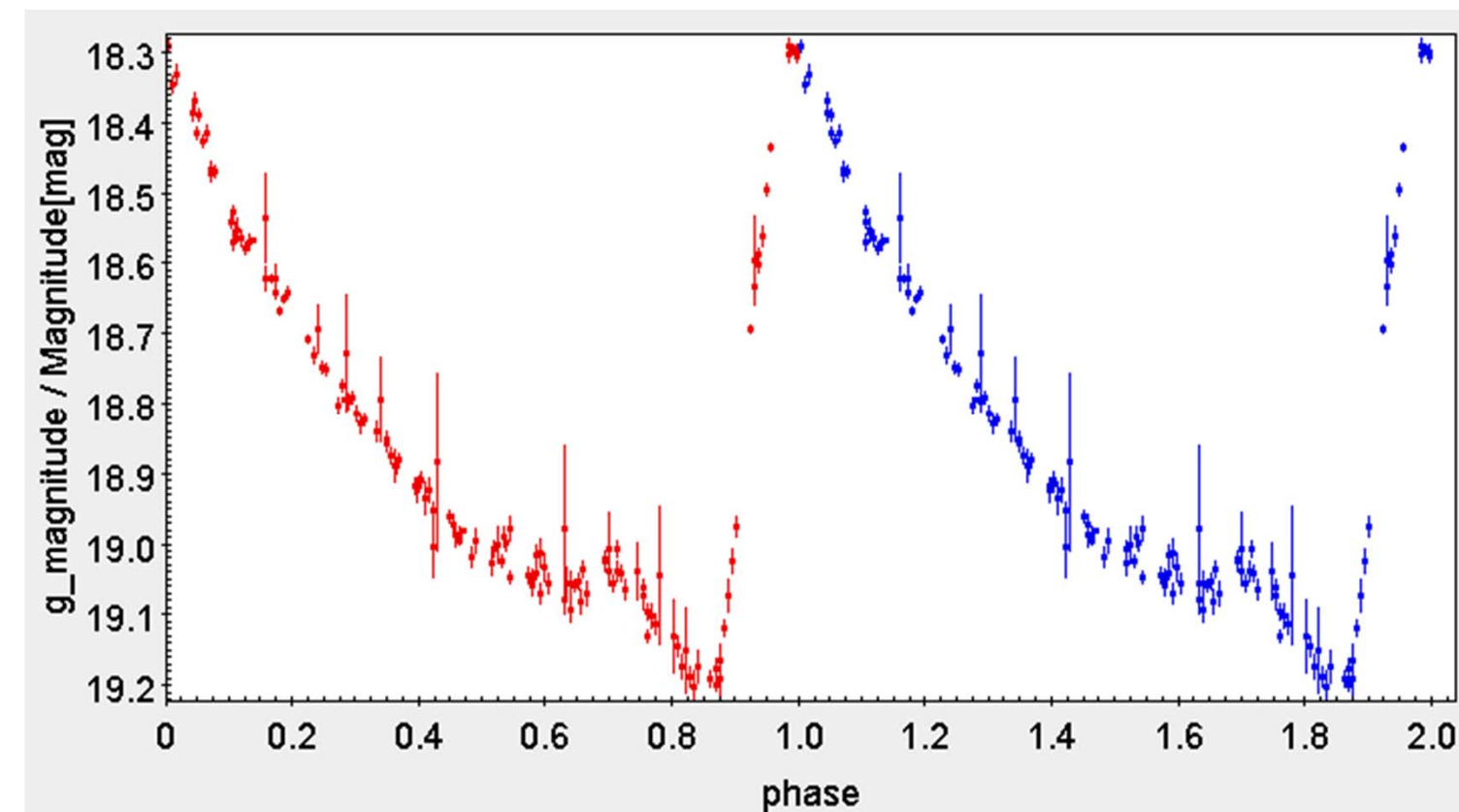
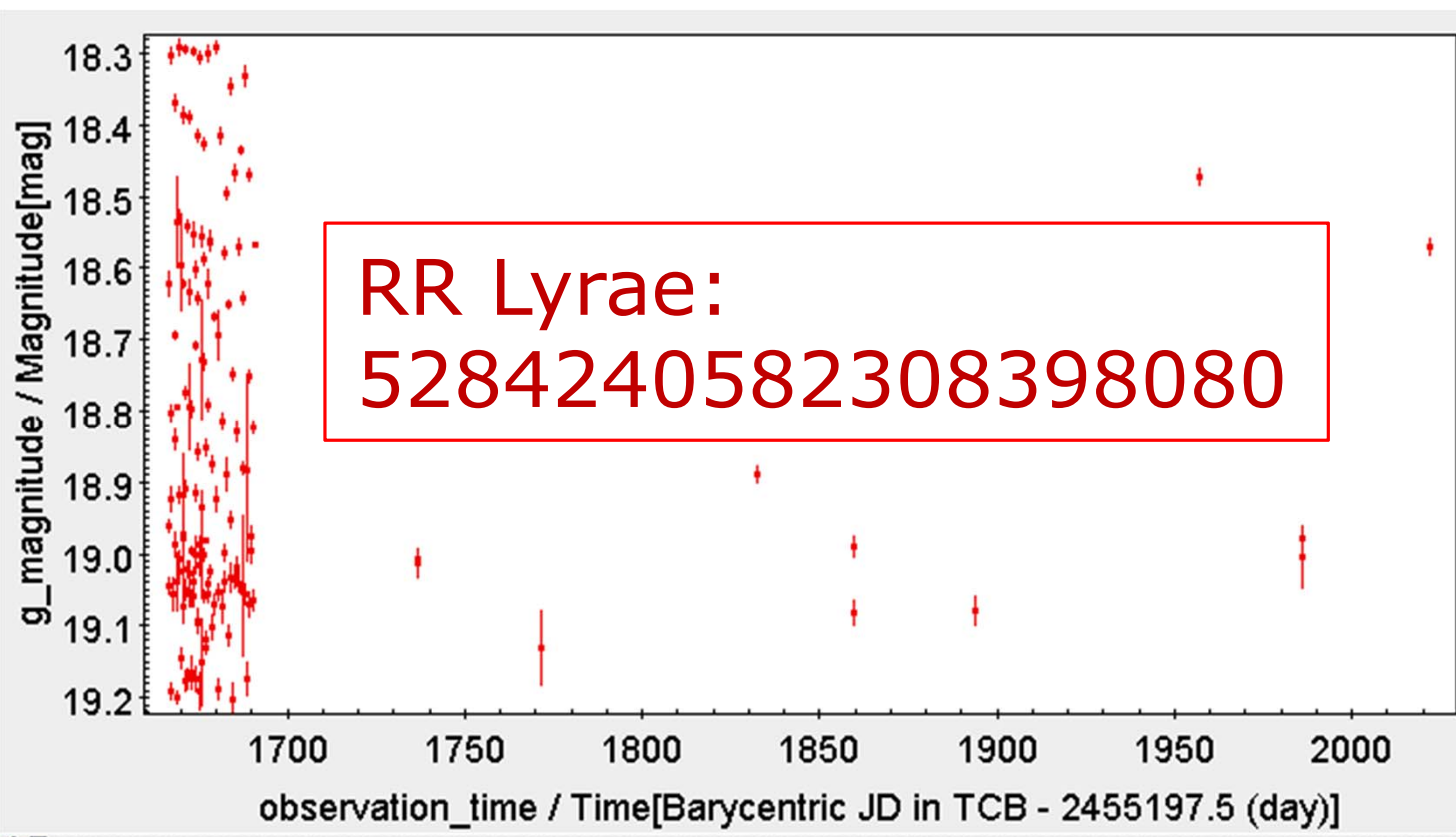
7. How to reconstruct an RR Lyrae light curve

Brown+ 2016: light curves

Plots are **reproducible!** ADQL queries in the paper

Uses Gaia epoch **photometry** and best fit **period**

Light curves in a **relational data base!** Best strategy? → DR2+



Brown+ 2016: light curves



select

```
curves.observation_time,
```

```
mod(curves.observation_time - rrlyrae.epoch_g, rrlyrae.p1)
```

```
  / rrlyrae.p1 as phase,
```

```
curves.g_magnitude,
```

```
2.5/log(10)* curves.g_flux_error/ curves.g_flux
```

```
  as g_magnitude_error
```

```
from gaiadr1.phot_variable_time_series_gfov as curves
```

```
inner join gaiadr1.rrlyrae as rrlyrae
```

```
  on rrlyrae.source_id = curves.source_id
```

```
where rrlyrae.source_id = 5284240582308398080
```


Brown+ 2016: light curves



select

```
curves.observation_time,  
mod(curves.observation_time - rrlyrae.epoch_g, rrlyrae.p1)  
    / rrlyrae.p1 as phase,  
curves.g_magnitude,  
2.5/log(10)* curves.g_flux_error/ curves.g_flux  
    as g_magnitude_error
```

Light curve reconstruction

```
from gaiadr1.phot_variable_time_series_gfov as curves  
inner join gaiadr1.rrlyrae as rrlyrae  
    on rrlyrae.source_id = curves.source_id
```

```
where rrlyrae.source_id = 5284240582308398080
```

Source selection

8. Additional resources

Partner data centres



Alternative and complementary **access** at data release time

Selected data. Custom interface. Additional functionality

AIP (Leibniz-Institut für Astrophysik Potsdam, Germany)

<https://gaia.aip.de/>

ARI (Astronomisches Rechen-Institut, Heidelberg, Germany)

<http://gaia.ari.uni-heidelberg.de/>

ASDC (ASI Science Data Center, Rome, Italy)

<http://gaiaportal.asdc.asi.it/>

CDS (Centre de Données astronomiques de Strasbourg, France)

<http://cdsweb.u-strasbg.fr/gaia>

Affiliated Data Centres. Receive full release after a small delay → efficient mirroring

Documentation and resources



Gaia Archive. Includes help and tutorials

<https://archives.esac.esa.int/gaia>

Gaia DR1 papers <http://www.cosmos.esa.int/web/gaia/dr1#A&A>

Online documentation (361 pages)

<http://gaia.esac.esa.int/documentation/GDR1/index.html>

Data model documentation

<https://gaia.esac.esa.int/documentation/GDR1/datamodel/>

ADQL: GAVO short course, UK ROE cookbook

<http://docs.g-vo.org/adql-gaia/html/index.html>

<https://gaia.ac.uk/science/gaia-data-release-1/adql-cookbook>

Gaia Helpdesk. <https://support.cosmos.esa.int/gaia/>

9. Conclusions

Conclusions



Archive: <https://archives.esac.esa.int/gaia>

Helpdesk: <https://support.cosmos.esa.int/gaia/>

Functionality: TAP+ (data base, user space, sharing), cross-match

Data: main tables, TGAS, variables, QSO, Xmatch, external catalogues

Bring code to the data: select, refine (ADQL, user tables) and download

Prepare for DR2 ($\sim 10^9$ sources). Archive might be the only way forward

Add **ADQL queries to your papers** (Brown+ 2016, Clementini+ 2016) → **Reproducibility**

Use it. User demand is a key driver for future developments

Ask us. Via Helpdesk for additional support. **Please provide feedback**

Want to support Gaia? → **keep writing papers** (and acknowledging)

ESA UNCLASSIFIED - For Official Use

A. Mora | The Gaia Archive | IAU Symposium 330 | 2017-04-26 | Slide 38

