Hands-on Session
Bringing together Gaia's astrometric and spectroscopic data

Jan Meichsner & Robin Geyer,
Lohrmann Observatory, TU Dresden

gaia

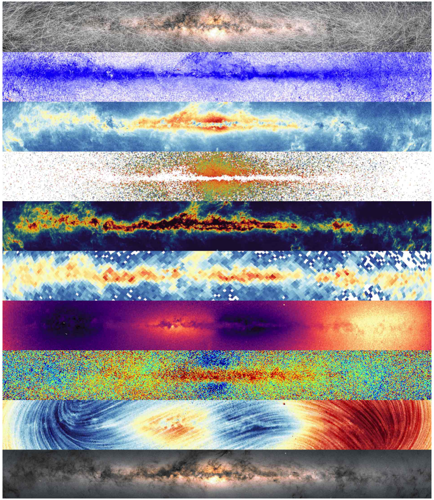
Gaia DPAC
Data Processing & Analysis Consortium

TECHNISCHE UNIVERSITÄT DRESDEN

2nd ChETEC-INFRA Observational School, Prague, July 2025

Outline of this hands-on session

- get you involved with Gaia data selected from the archive
- working with TOPCAT
- focus on the catalogues of DR3 with astrometry, astrophysical parameters, and spectra
- this slides can be obtained from `https://lohrmobs.de/HandsOnGaia`



Composition of Gaia sky maps, based on different Gaia data products from Gaia DR3.

Source: ESA/Gaia/DPAC

Who are we

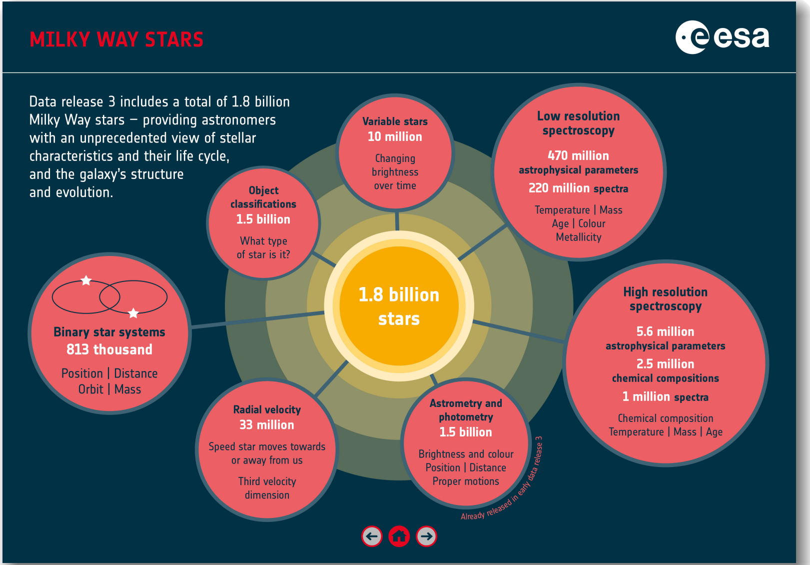
Robin Geyer

- studied computer science at the Technische Universität Dresden
- worked/studied in HPC/supercomputing
- since 2015 member of Gaia DPAC at TUD
- main tasks:
  - Gaia solution on HPC
  - gravitational waves and astrometry
  - analyse data

Jan Meichsner

- studied physics at the Technische Universität Dresden
- PhD in Mathematics at the Technische Universität Hamburg
- since 2022 member of Gaia DPAC at TUD
- main tasks:
  - be the math support
  - modelling the basic angle variations
  - analyse data

Gaia Data Products I



Gaia Data Products II

- Simple data ("plain tables" with unique `source_id`)
  - Mean astrometry, photometry, radial velocity, classifications, etc.
  - Astrophysical parameters from spectras (BP/RP, RVS)
  - Collections of special sources: extra-galactic, reference frame, non-single stars, solar system observations, ...
- Large or complex ("ancillary") data
  - Epoch photometry (light curves)
  - Mean BP/RP spectra
  - Mean RVS spectra
  - Classification data
  - ...

Gaia Data Products III

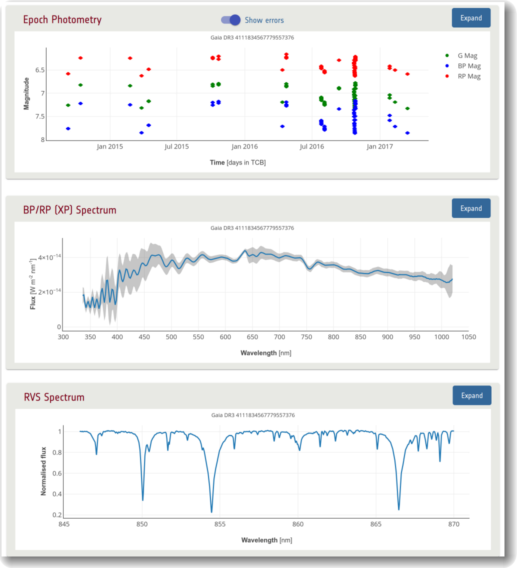
An example of tabular data

key	value	key	value
designation	Gaia DR3 4111834567779557376	has_xp_continuous	true
source_id	4111834567779557376	has_xp_sampled	true
ra	256.5229102004341	has_rvs	true
dec	-26.580565130784702	has_epoch_photometry	true
parallax	1.1537674647874 78	has_epoch_rv	true
pmra	0.3895503169963796	has_mcmc_gspphot	true
pmdec	-0.2892653937514547 7	has_mcmc_msc	true
ruwe	0.83691496	teff_gspphot	5934.4375
phot_g_mean_mag	7.069333	logg_gspphot	1.769
bp_rp	1.1675663	mh_gspphot	-0.5115
radial_velocity	-27.150185	distance_gspphot	1024.4829
phot_variable_flag	VARIABLE	azero_gspphot	1.0671
ebpminrp_gspphot	0.4736	ag_gspphot	0.8745
non_single_star	0		

Gaia Data Products IV

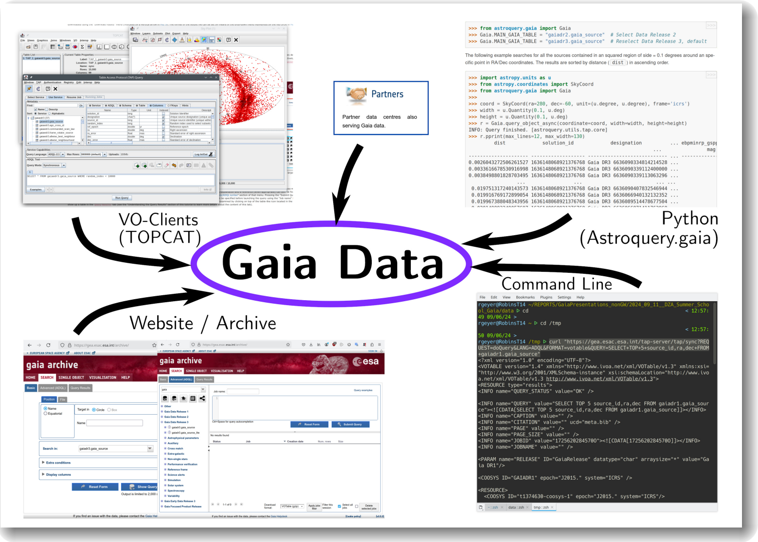
An example of large ancillary data. More exist:

- Markov-Chain Monte-Carlo samples for classification
- Specific Object Studies
- All-sky total galactic extinction maps
- ...



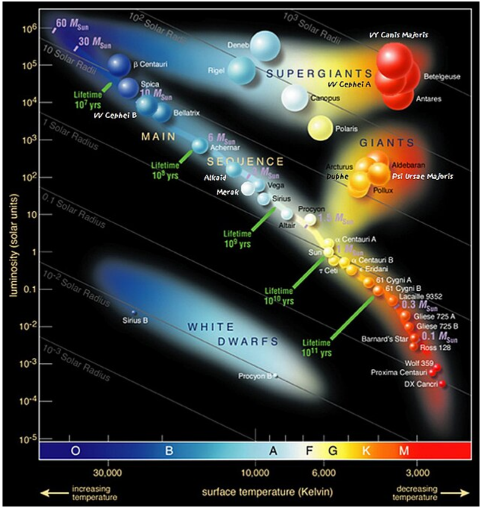
Data Access

There are multiple ways to access the data products



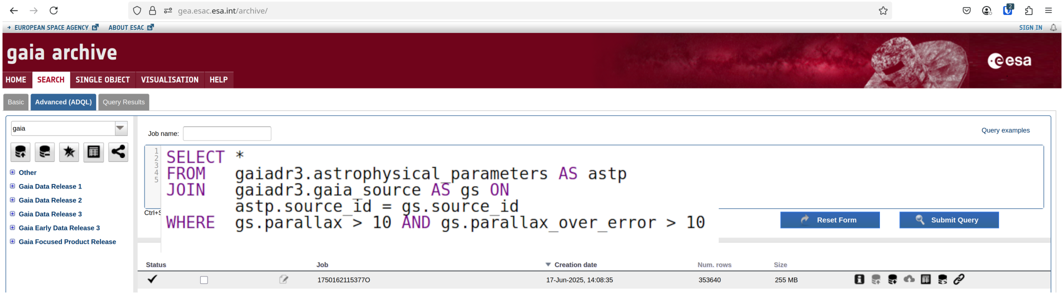
The Hertzsprung-Russell diagram

- created 1911 by Ejnar Hertzsprung and in 1913 by Henry Norris Russell
- shows the relation between luminosity of a star and its surface temperature
- the difference between apparent brightness in two spectral bands (`BP - RP`), commonly called *colour* in astronomy, can be used as proxy for temperature



Source: https://commons.wikimedia.org/wiki/File:Updated_Hertzsprung-Russell_Diagram.jpg, Autor: Daniel William “Danny” Wilson, License: CC BY-SA 4.0

Part I: Querying the Gaia Archive

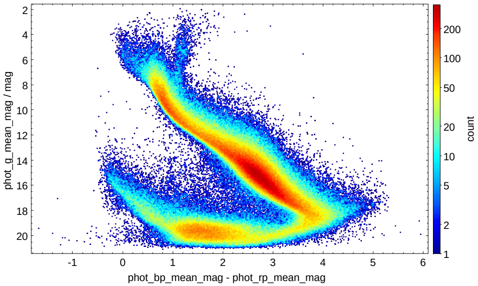


- use the Gaia archive (https://gea.esac.esa.int/archive/)
  - click *SEARCH*, *Advanced (ADQL)*
  - use the query from the screenshot
  - left-hand side tells you about available tables
  - join smaller on bigger table
  - also check out *HELP*

The query give **353.640 individual sources**. Save the dataset to your local disk.

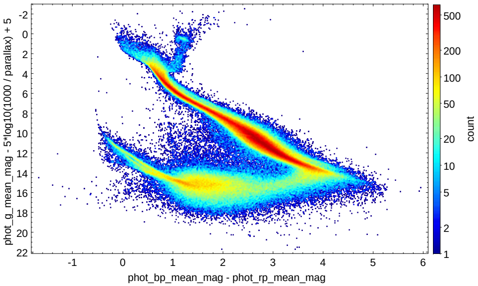
Part II: HR diagram

- use the Graphics | Plane Plot menu item or toolbar button
  - use , to load the dataset you just downloaded
  - **X** `"phot_bp_mean_mag - phot_rp_mean_mag"`
  - **Y** `"phot_g_mean_mag"`
  - use **Axes** | **Coords** | **Y Flip**
  - change under **Form** the point **Mode** to **weighted** and **Combine** to **count**
  - use further **Aux Axis** | **Shader** | **Rainbow2**
  - tick **Shader Flip**
  - choose **Scaling: log**



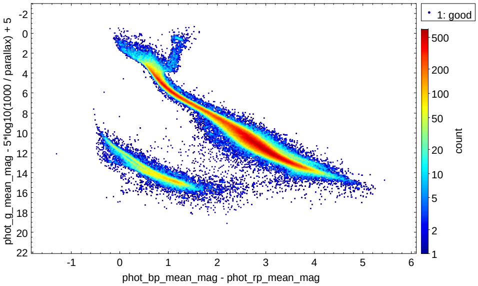
HR diagram II

- The HRD from the dataset looks noisy
- using the **absolute magnitude** in the Y axis gives some improvement:
  Y `"phot_g_mean_mag - 5*log10(1000 / parallax) + 5"`



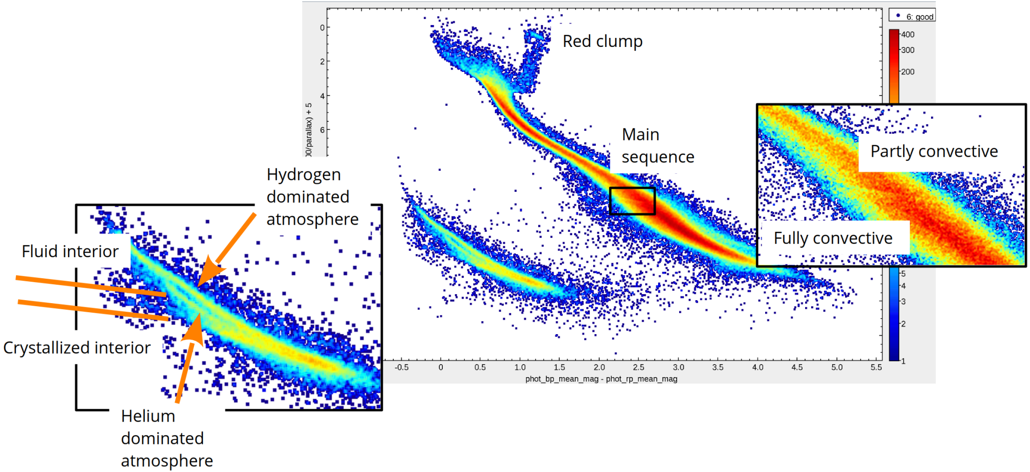
HR diagram II

- The HRD from the dataset looks noisy
- using the **absolute magnitude** in the Y axis gives some improvement:
  Y `"phot_g_mean_mag - 5*log10(1000 / parallax) + 5"`
- apply some quality filtering on the data.
  - use (**Display row subsets**) and create ( ) the subsets:
  - **goodRuwe:** `ruwe` $<$ 1.2
  - **goodColour:** `phot_rp_mean_flux_over_error` $>$ 10 && `phot_bp_mean_flux_over_error` $>$ 10
  - **good:** `goodRuwe && goodColour`



HR diagram III

A lot of astrophysics already visible here ...

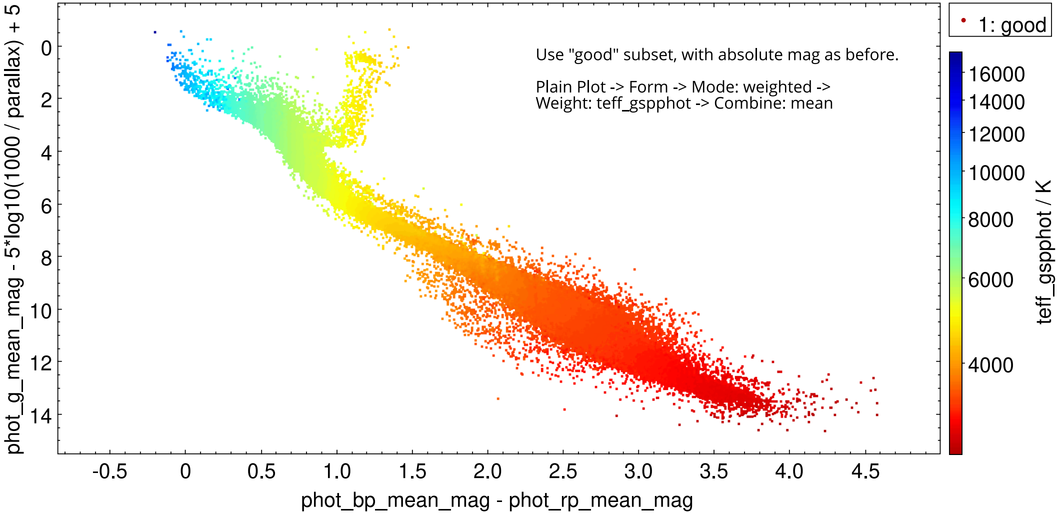


"Jao gap"
Jao et al. (2018)

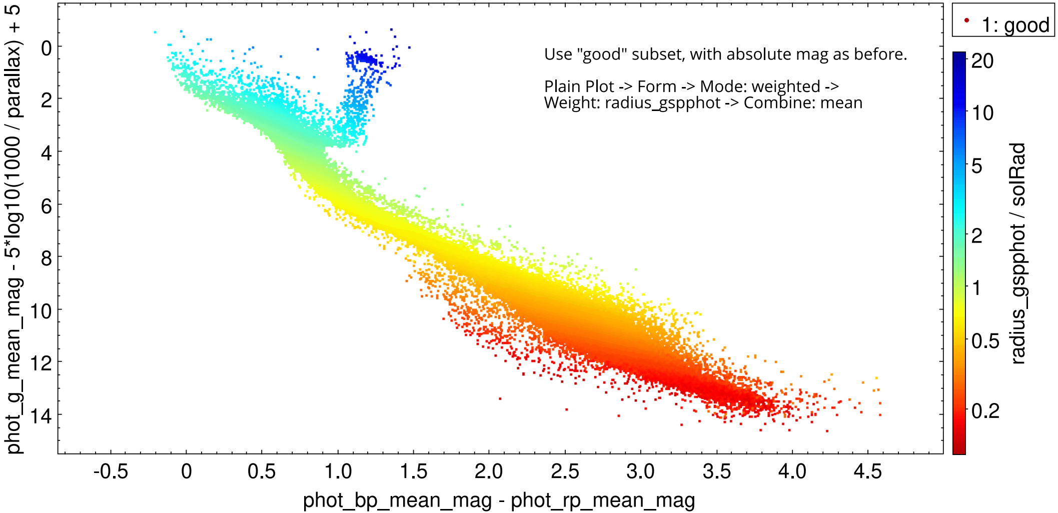
Part III: And now you - playing time

- find Columns telling you something about temperature T (search for "`teff`") and define sub-selections, say, $T < 5000\,\mathrm{K}$, $5000 < T < 10000\,\mathrm{K}$, and $T < 10000\,\mathrm{K}$. Make them visible in the HR diagram
- same for star radii R ($R < 0.5\,\mathrm{R_{Sun}}$, $0.5\,\mathrm{R_{Sun}} < R < 1.0\,\mathrm{R_{Sun}}$, $R > 1.0\,\mathrm{R_{Sun}}$)
- find a smarter solution to the first two points than defining discrete categories (Hint: Play with the `Weight` option of the shader)

The main sequence when you colour code with temperatures



The main sequence when you colour code with radii



Spectra where?

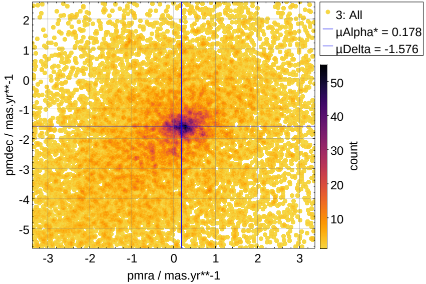
- define **goodSpec:** `good && has_xp_sampled == true`
- have a look on `Display table cell data` and highlight this subset. Grab any source ID from there.
- go back to the Gaia archive and try this source ID under **SINGLE OBJECT** as `Gaia DR3 YOUR_SOURCEID`
- check out the spectrum

Part IV: And what if I have a name but no source IDs?

- we want to have a look at the *Eagle Nebula* (M16), in particular the “Pillars of Creation”
- go back to the Archive (`https://gea.esac.esa.int/archive/`) and use *Basic* (tick *Name*) to search for M16 with a radius of 25 arcmin
- before submitting the query, use *Show Query*, and replace everything between `SELECT` and `FROM gaiadr3.gaia_source` by $*$. Also get rid of `ORDER BY target_id, "target_separation (deg)"` at the end
- download the result and load the table in TOPCAT

Proper motion space

- try to separate Eagle Nebula members from foreground/background sources.
- plot sources in proper motion space
  - use the Graphics — **Plane Plot** menu item or toolbar button
  - **X** `"pmra"`
  - **Y** `"pmdec"`
  - adjust the plotted points under **Form** using
    - **Mode** weighted, **Combine** count
- we know from SIMBAD that the PM of the cluster should be $\mu_{\alpha*} = 0.178\,\text{mas/yr}$ and $\mu_\delta = -1.576\,\text{mas/yr}$
- graphically select this co-moving cluster as new subset
  - select the blob using button, drag mouse, again
  - **New Subset Name** `"pavlo_sel"`



Additional stuff

Data Access – Links

Other ways to access the data:

- TOPCAT can be downloaded here
  - https://www.star.bris.ac.uk/~mbt/topcat/
- how-to for the command line API can be found here
  - https://www.cosmos.esa.int/web/gaia-users/archive/programmatic-access#CommandLine_Tap
- how-to for access via Python
  - astroquery.gaia documentation: https://astroquery.readthedocs.io/en/latest/gaia/gaia.html
  - simple Python access: https://www.cosmos.esa.int/web/gaia-users/archive/use-cases#ClusterAnalysisPythonTutorial
  - complex Python access: https://www.cosmos.esa.int/web/gaia-users/archive/datalink-products#datalink_jntb_get_all
- a list of partner data centres can be found under
  - https://www.cosmos.esa.int/web/gaia/data-access#PartnerDataCentres

Data Model Descriptions

When writing ADQL queries, or searching for something specific, its good to know what data exist.

- a list of all tables and their content can be found under
  - for DR3: https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/
  - for FPR: https://gea.esac.esa.int/archive/documentation/FPR/index.html
- the ADQL query website and TOPCAT also have short descriptions for each field

ADQL Tips and Tricks I

- ADQL warnings: POWER, SIN, COS are slow and 1/3600 is not 1./3600.
- Use gaia_source_lite: fewer (specialised) columns give faster queries
- Don't be lazy: don't SELECT *, retrieve only your fields of interest
- Use indexed fields + pre-computed fields + constraints, see bold fields in the web GUI tooltips for indexed fields and the data model for pre-computed fields (e.g., bp_rp and parallax_over_error)
- Optimise cross-matches: place the smallest catalogue first and use JOIN + ON
- Avoid TOP + ORDER BY: use two queries and JOB_UPLOAD to sort the result
- Randomise use the random_index field to define representative samples
- Divide and conquer: split big queries in chunks using random_index
- Ask the Helpdesk: ask for extended time-out or storage quota or query advise

A funny task for fast students: explain what the “objects” in this proper-motion scatterplot are, or more general how you can overall explain this plot. This is from a 1 deg cone search around M16.