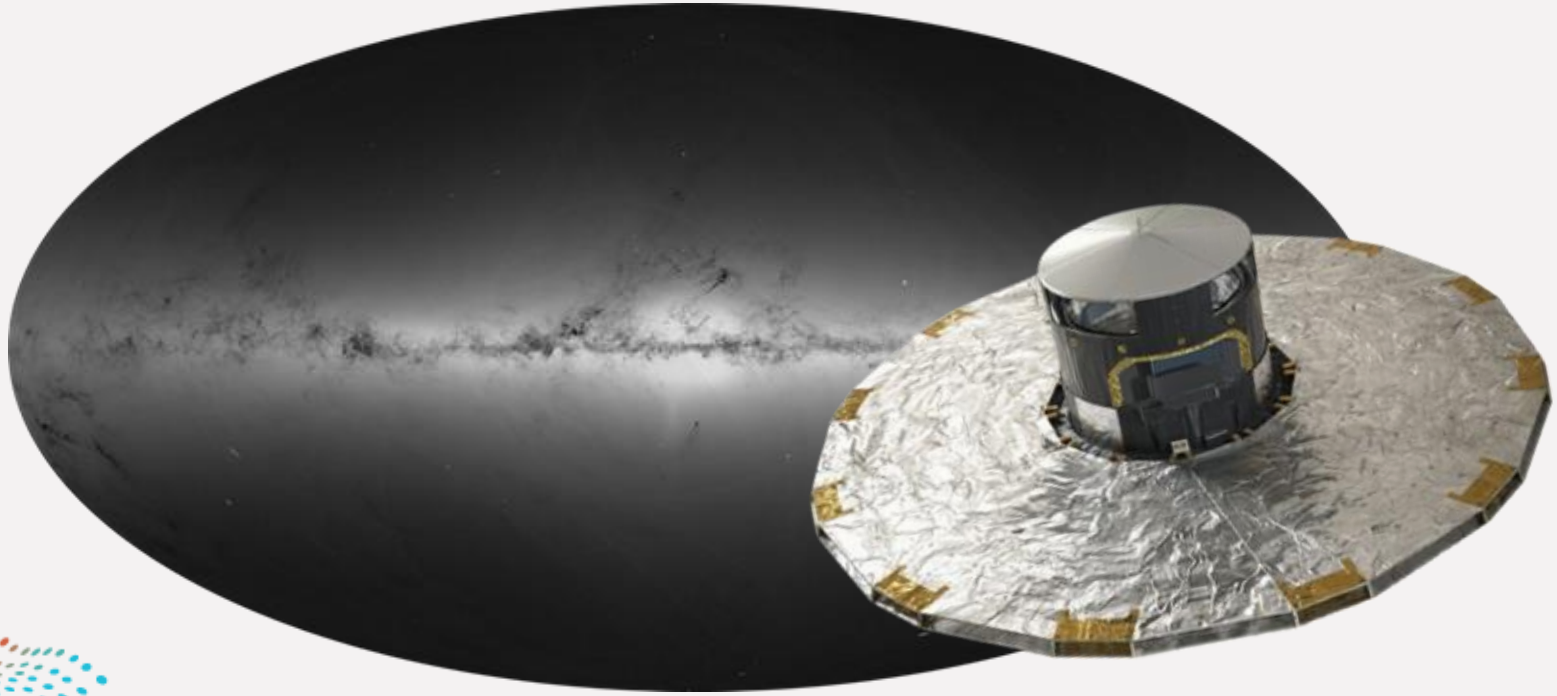
Gaia Hackathon 2026



MIT KAVLI INSTITUTE

Prof. Lina Necib's group

COSMIC CODEBREAKERS

Streams

Elliot Davies, PhD

Nathaniel Starkman, PhD

Brinson Fellow

Galactic Center/Halo

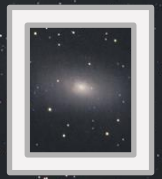
Abdelaziz Hussein

Gonzalo Herrera, PhD
NTN Fellow

Xiaowei Ou, PhD' 25

After MIT:
GECO + CosmicAI Fellow, UVA

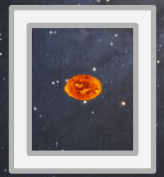
Dwarf Galaxies



Tri Nguyen, PhD' 24

After MIT:
CIERA Fellow,

Solar Neighborhood



Zeineb Mezghanni

Xiuyuan Zhang

Lina Necib, MIT

Schedule for the Week

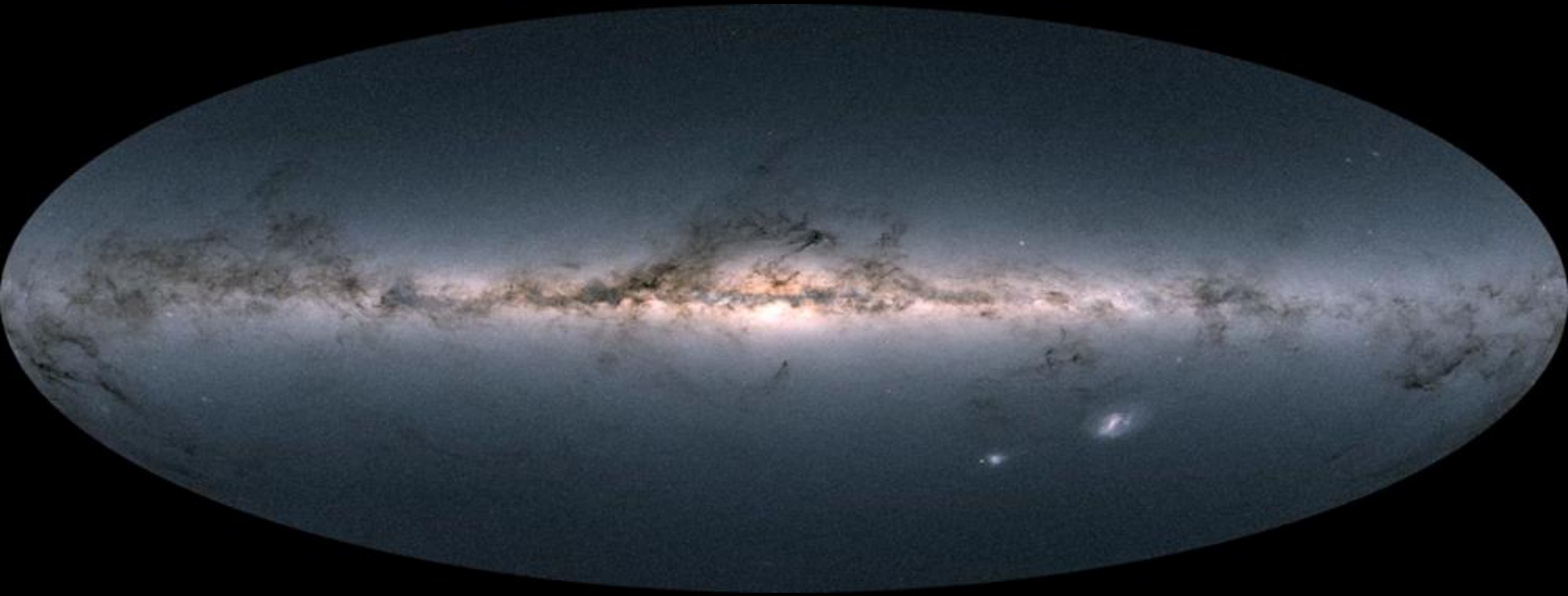
Time	Topic
Thurs Jan 29	Data Bootcamp: - Research Practices - Working with Gaia data Hackathon! (afternoon)
Fri Jan 30	Hackathon!
Sat Jan 31	Hackathon!

- Research Project Setup
- Intro to Gaia
- Running Jobs on the Cluster
- Coordinate Transforms
- Actions & Angles
- TOPCAT

Gaia

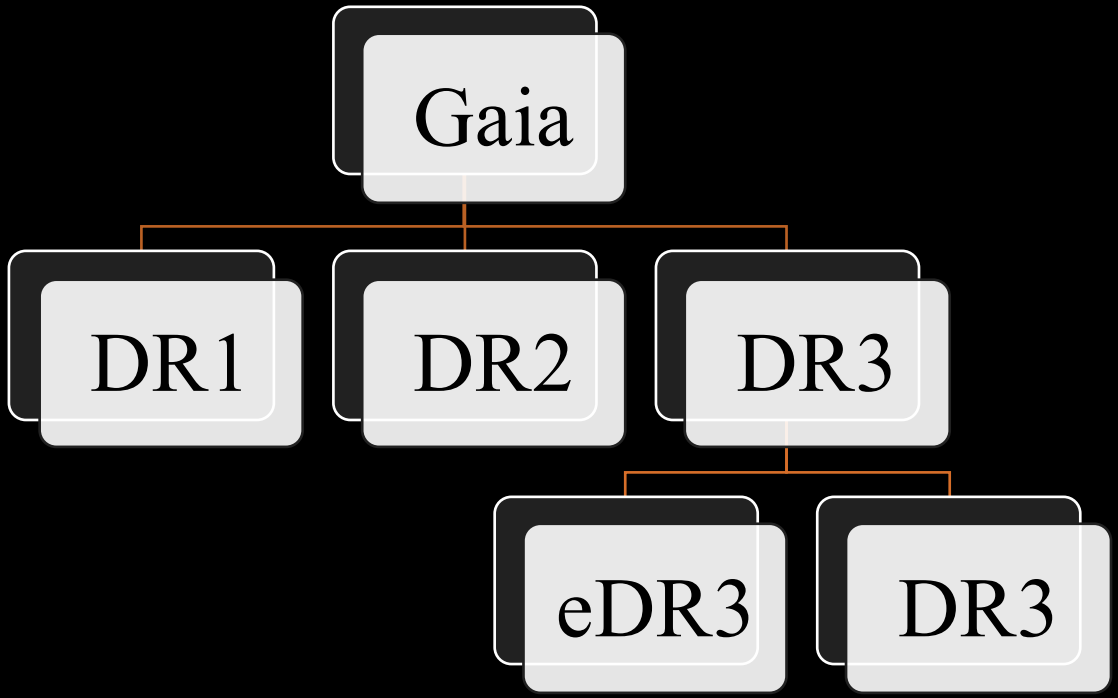
- December 2013 – January 2025
- **Data:** Positions and perpendicular velocities (5D Kinematics) of 1.5 billion stars.
- **Second Data Release:** Line-of-sight velocities (6D Kinematics) of 7 million stars.
- **Third Data Release:** Line-of-sight velocities (6D Kinematics) of 33 millions stars.

Gaia



Credit: Gaia Sky; S. Jordan / T. Sagristà

Details of the Gaia data



Gaia DR3 in numbers

	# sources in Gaia DR3	# sources in Gaia DR2	# sources in Gaia DR1
Total number of sources	**1,811,709,771**	**1,692,919,135**	**1,142,679,769**
	Gaia Early Data Release 3		
Number of sources with full astrometry	1,467,744,818	1,331,909,727	2,057,050
Number of 5-parameter sources	585,416,709		
Number of 6-parameter sources	882,328,109		
Number of 2-parameter sources	343,964,953	361,009,408	1,140,622,719
Gaia-CRF sources	1,614,173	556,869	2191
Sources with mean G magnitude	1,806,254,432	1,692,919,135	1,142,679,769
Sources with mean G_{BP}-band photometry	1,542,033,472	1,381,964,755	-
Sources with mean G_{RP}-band photometry	1,554,997,939	1,383,551,713	-
	New in Gaia Data Release 3	Gaia DR2	Gaia DR1
Sources with radial velocities	33,812,183	7,224,631	-

Gaia DR3 in numbers

5D Kinematics:
- 2 Angular positions in the sky
- 2 velocities in the perpendicular plane
- Distance to the star/parallax

	# sources in Gaia DR3	# sources in Gaia DR2	# sources in Gaia DR1
Total number of sources	**1,811,709,771**	**1,692,919,135**	**1,142,679,769**
	Gaia Early Data Release 3		
Number of sources with full astrometry	1,467,744,818	1,331,909,727	2,057,050
Number of 5-parameter sources	585,416,709		
Number of 6-parameter sources	882,328,109		
Number of 2-parameter sources	343,964,953	361,009,408	1,140,622,719
Gaia-CRF sources	1,614,173	556,869	2191
Sources with mean G magnitude	1,806,254,432	1,692,919,135	1,142,679,769
Sources with mean G_{BP}-band photometry	1,542,033,472	1,381,964,755	-
Sources with mean G_{RP}-band photometry	1,554,997,939	1,383,551,713	-
	New in Gaia Data Release 3	Gaia DR2	Gaia DR1
Sources with radial velocities	33,812,183	7,224,631	-

Gaia DR3 in numbers

5D Kinematics:
- 2 Angular positions in the sky
- 2 velocities in the perpendicular plane
- Distance to the star/parallax

	# sources in Gaia DR3	# sources in Gaia DR2	# sources in Gaia DR1
Total number of sources	**1,811,709,771**	**1,692,919,135**	**1,142,679,769**
	Gaia Early Data Release 3		
Number of sources with full astrometry	1,467,744,818	1,331,909,727	2,057,050
Number of 5-parameter sources	585,416,709		
Number of 6-parameter sources	882,328,109		
Number of 2-parameter sources	343,964,953	361,009,408	1,140,622,719
Gaia-CRF sources	1,614,173	556,869	2191
Sources with mean G magnitude	1,806,254,432	1,692,919,135	1,142,679,769
Sources with mean G_{BP}-band photometry	1,542,033,472	1,381,964,755	-
Sources with mean G_{RP}-band photometry	1,554,997,939	1,383,551,713	-
	New in Gaia Data Release 3	Gaia DR2	Gaia DR1
Sources with radial velocities	33,812,183	7,224,631	-

6D Kinematics:
- Adding line-of-sight velocities

What's a Hackathon?

- A collaborative environment for us to all work on the same data, but doing our own things.
- A bunch of people coding up in the same place to just have some fun!
- Logistics:
  - There will be lunch served everyday!
  - There will be snacks too!
  - We will have (completely voluntary) presentations at 9:30am and 4pm on Friday and Saturday. [Add a figure to a common Google doc. Whoever wants to speak, will talk for at most 2 minutes/1 slide.]

What's a Hackathon?

Goals

- Download the data
- Pre-process into formats we can use
- Play with it and discover some fun things!

One of the most important things we will cover is ``Cluster Etiquette''

We will be coding on SubMIT!

The data is already there, so we will all be able to access it (without copying over and over multiple TB of data!)

However, we will learn:

1. Don't run anything heavy on the login node (unless you want to get grumpy emails/get kicked out. It is easy to find out who you are).

    This also applies for jupyter notebooks and VS Code!

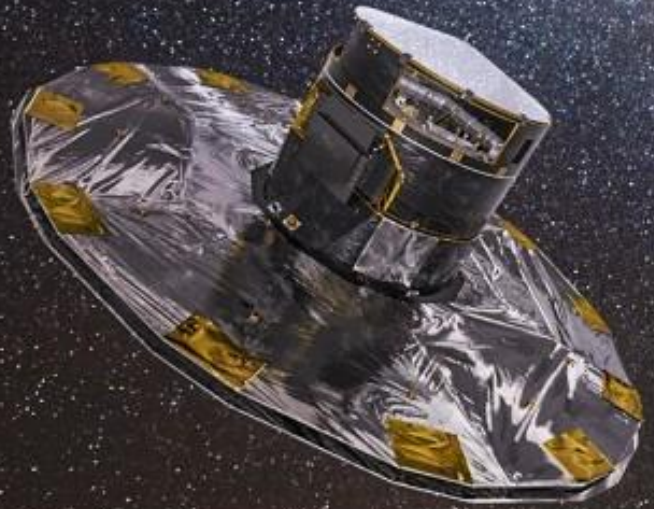
2. How to submit a job.

3. Every cluster is a little special, but you need to be able to run the same code. Things like virtual env/anaconda will be truly helpful.

Prizes:

- There will be three $100 Amazon gift cards for:
  - Coolest Plot
  - Cleanest Code
  - Best Science Result
- To qualify:
  - You have to present whatever prize you want to be considered for at one of the presentations during the hackathon.
- Prizes will be announced Saturday at 4:30pm.

I hope to see some of you there!



Gaia Hackathon