

Gaia Core Processing on IRIS

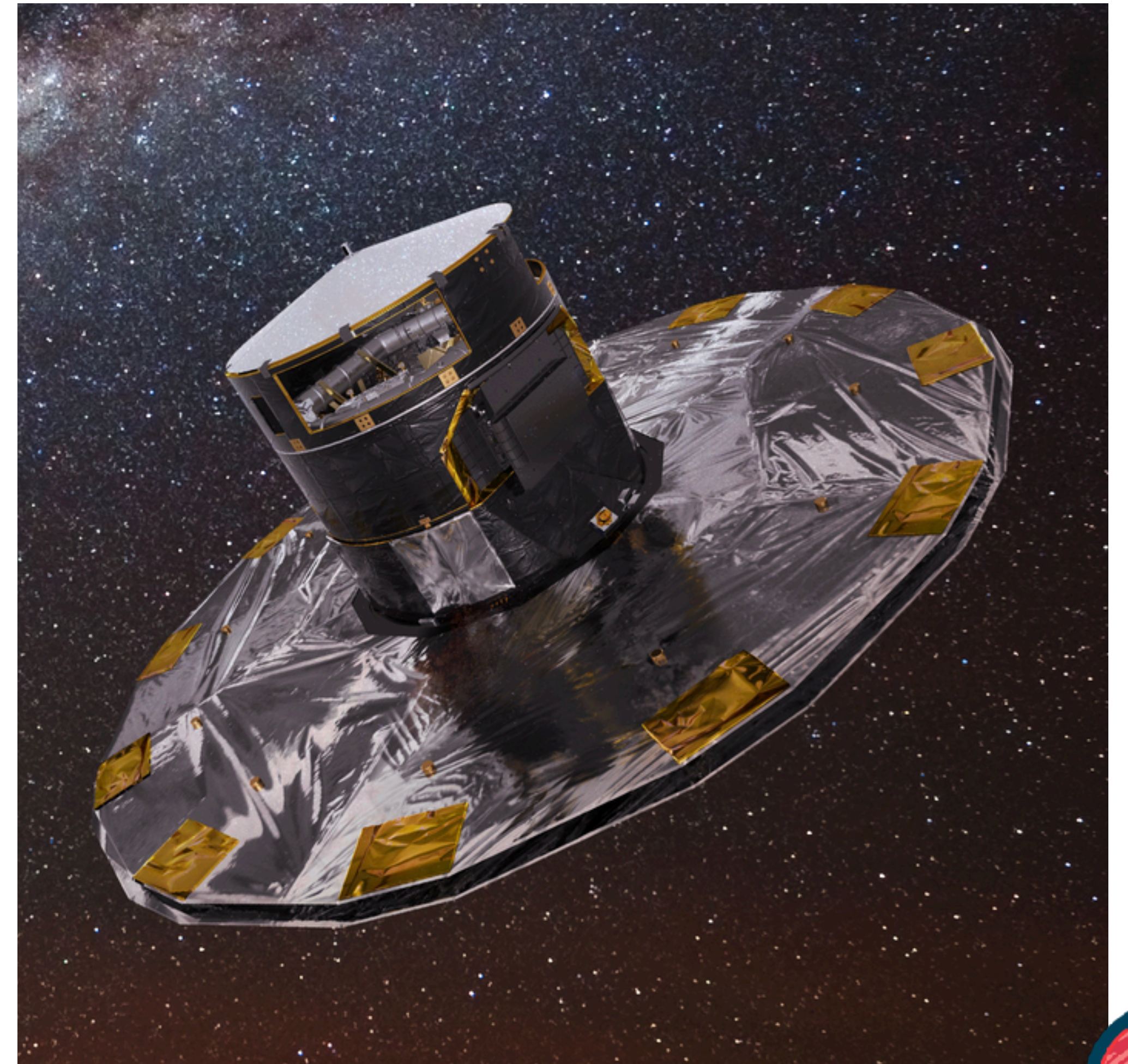
Patrick Burgess

- Gaia mission.
- CU5 group and Cambridge data processing.
- IRIS allocation at Cambridge.
- What/how we use this currently.
- 2025 and beyond.



Gaia mission

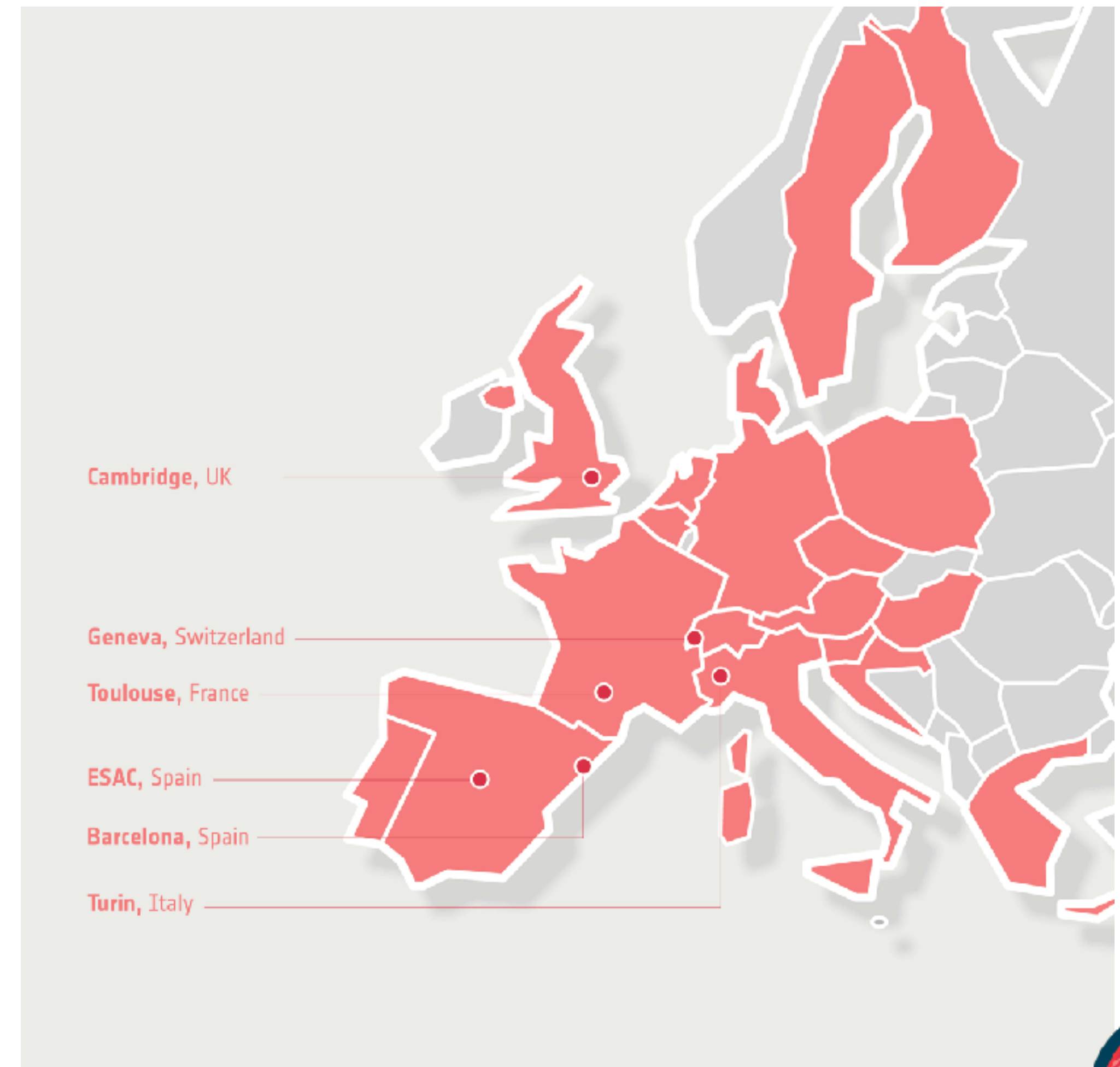
- European Space Agency astrometry mission building the most precise three dimensional map of our galaxy.
- https://www.esa.int/Science_Exploration/Space_Science/Gaia_overview
- Gaia satellite launched in December 2013
- Observations started in July 2014 and final science observations are taken in January 2025. Satellite is continuously observing.
- Final data processing taking place over several years.
- 265,782,096,362 FOV observations so far.



DPAC

Data Processing and Analysis Consortium

- 450 scientists and engineers from over 20 countries.
- Contributions from more than 100 different research institutions.
- Cambridge is one of the 6 data processing centres.
- Data Releases:
 - DR1: September 2016
 - DR2: April 2018
 - EDR3: December 2020
 - DR3: June 2022
 - DR4: in progress
 - Cambridge finishing this now.
 - DR5: In progress



Cambridge Processing

- CU5 group: Photometric processing within DPAC. DPCI - Data Processing Center at Institute of Astronomy, Cambridge.
- Cyclic processing: batch processing of entire mission so far, carried out once per data release.
 - Based on Apache Spark data processing framework. Internally developed software implemented in Java and Python.
 - Photometric Core Processing, called PhotPipe, handling photometry and low resolution spectra.
 - Generate calibrations to remove observational effects.
 - Generate mean and epoch data types for publication.
 - Source Environment Analysis, called SEAPipe
 - Analyse the combined observations of each source to determine the source environment - undetected companion sources, extended sources - and generate data for publication.
- Photometric Science Alerts: <https://ioadpc.ast.cam.ac.uk/alerts>, called AlertPipe
 - Analyse daily data and provide rapid notification of transient events to enable ground based followup. 26,735 currently published.

Cambridge Gaia Systems

- Several iterations of Gaia data processing resources at Cambridge.
- 2008: Dedicated self managed, 12 servers.
- 2012: Dedicated self managed, 130 servers, 2PB disk storage.
- 2016: Dedicated data centre managed, 250 servers, 4PB disk storage.
- 2024: Transfer almost entirely to IRIS cloud resources. Previous system mostly decommissioned in April 2024.



IRIS Resources

- Hosted in West Cambridge Data Center, local university run facility. Provided as an allocation of Openstack resources on the Arcus cloud.

	vCPU	Disk TB
2023-2024	12,848	4,272
2024-2025	12,336	4021

Reduction between 2023/24 and 24/25 due to decision taken to extend support for some legacy hardware (AlertPipe)

- Tied to specific hardware: Dual Intel Xeon Icelake 8368Q, total 148 vCPU, 480GiB RAM available for instances. 3.2TiB local storage. 100GbE.

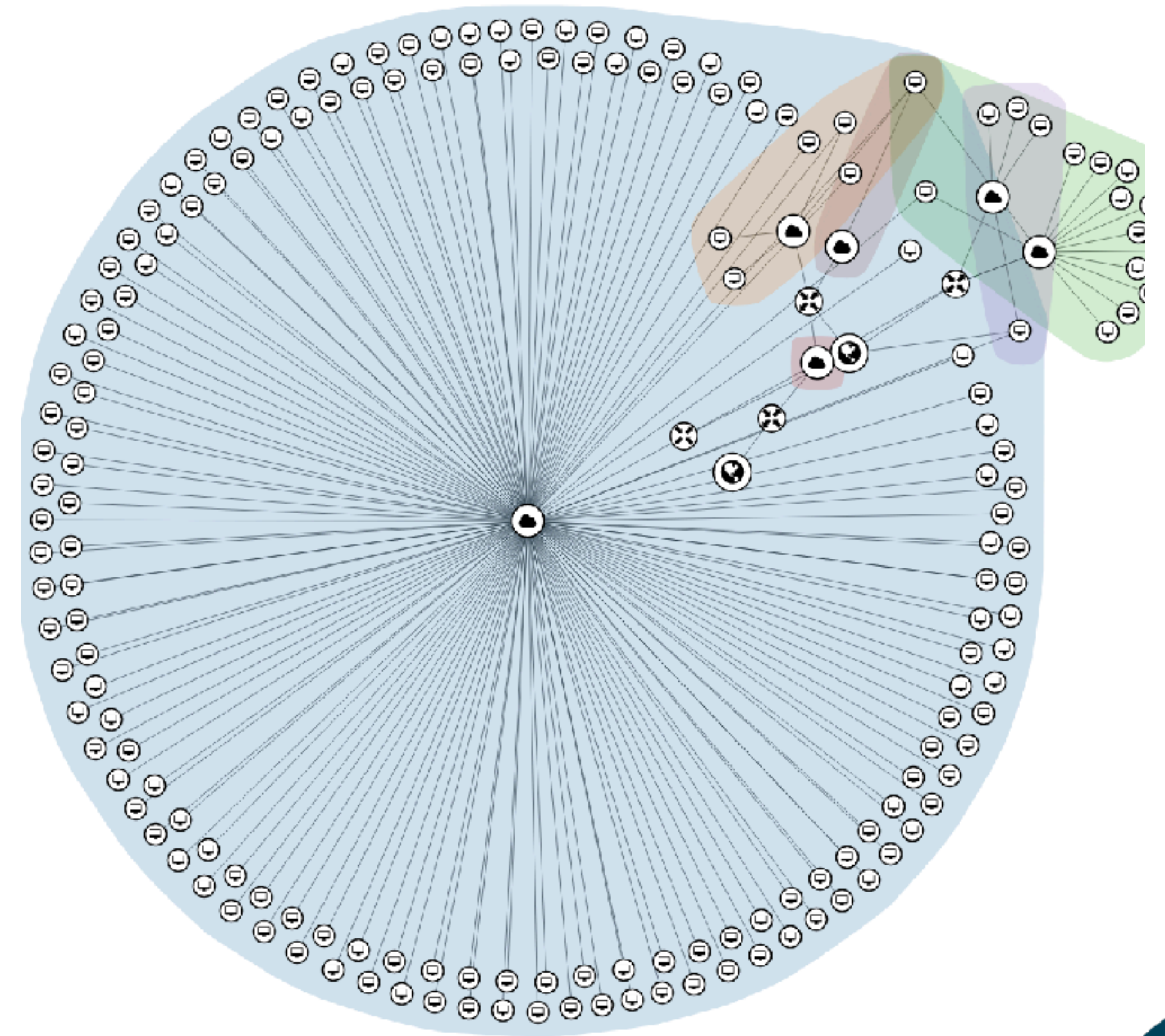
Usage

IRIS Resources provide a variety of functions to the project:

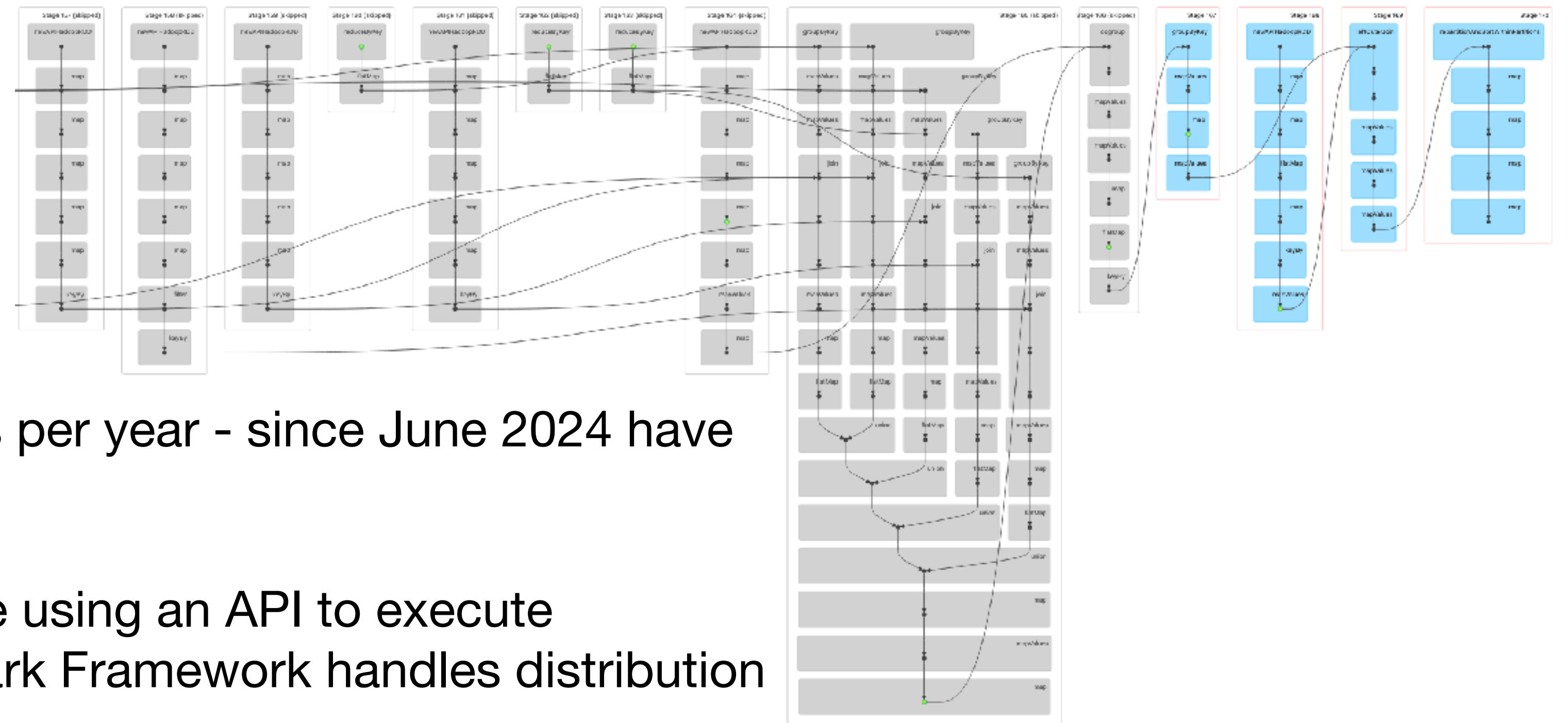
- Processing Cluster(s). The majority of resources.
- Webserver.
- Data transfers via Aspera. Data management software.
- Continuous integration, Jenkins.
- Software quality, SonarQube.
- Nexus software artefact repository.
- Validation database and query software.
- Resources for off cluster ad-hoc analysis.
- Systems infrastructure, monitoring, OS updates, logging.

Openstack and Ansible

- All IRIS resources managed via Openstack. Using web UI, command line or Ansible.
- All Gaia Core Processing infrastructure defined, deployed and managed via Ansible code.
 - Approximately 12k lines of configuration including yaml, j2, xml, etc.
- Previous systems mostly manually configured with some use of puppet.
- Current system does not enforce compliance, relies on good practice.



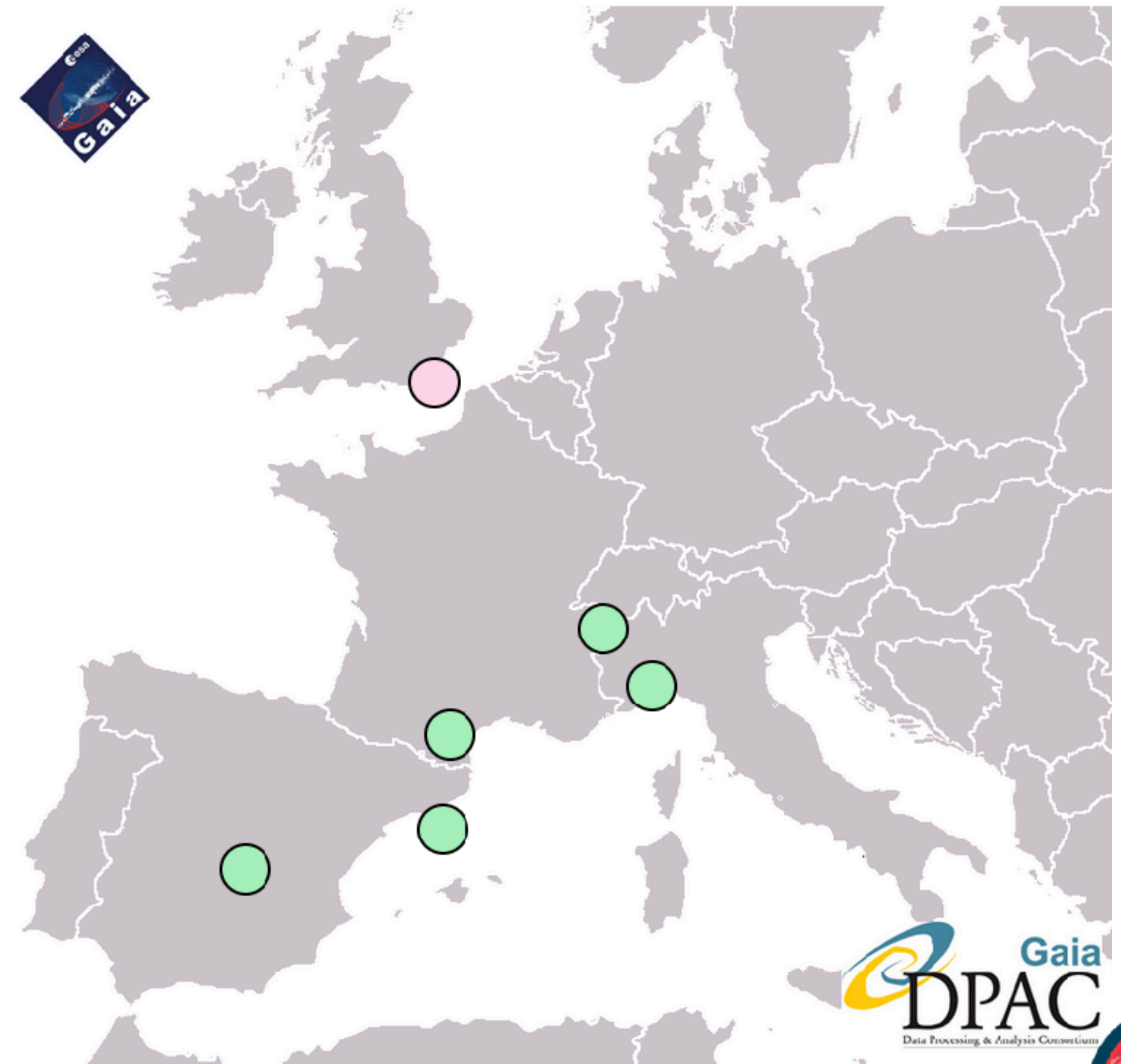
Apache Spark



- Execute thousands of Spark jobs per year - since June 2024 have 4,635 recorded.
- Define a Spark flow as Java code using an API to execute transformations and actions. Spark Framework handles distribution of processing over the cluster.
- Currently executing on a Standalone Spark cluster. Jobs are submitted to the Spark Master process and distributed to Worker nodes.

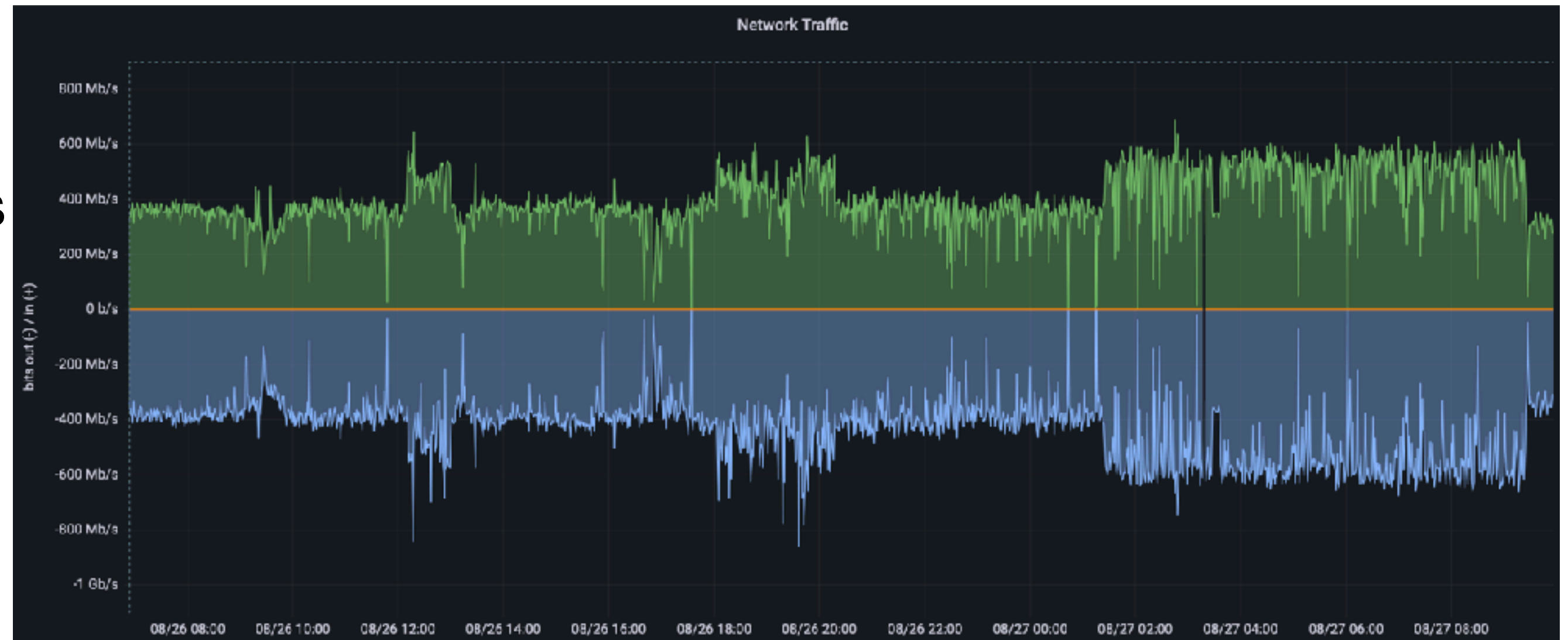
Aspera

- All operational data exchange is via Aspera.
- <https://www.ibm.com/products/aspera>
- Licensed software supplied by ESA to coordinate secure and reliable data transfer between data processing centres.
- Using a hub and spokes model. All Data Processing Centres communicate with Madrid.
- Allows all transfers to be tracked and backed up.
- Approx. 250,000 transfers, 13.5 million files incoming to Cambridge so far.



Transfer performance

- Transfer performance is significant. Outgoing and incoming cyclic datasets reach 100-200TB scale.
- Previously achieving approx. 1.8Gb/s sustained transfers.



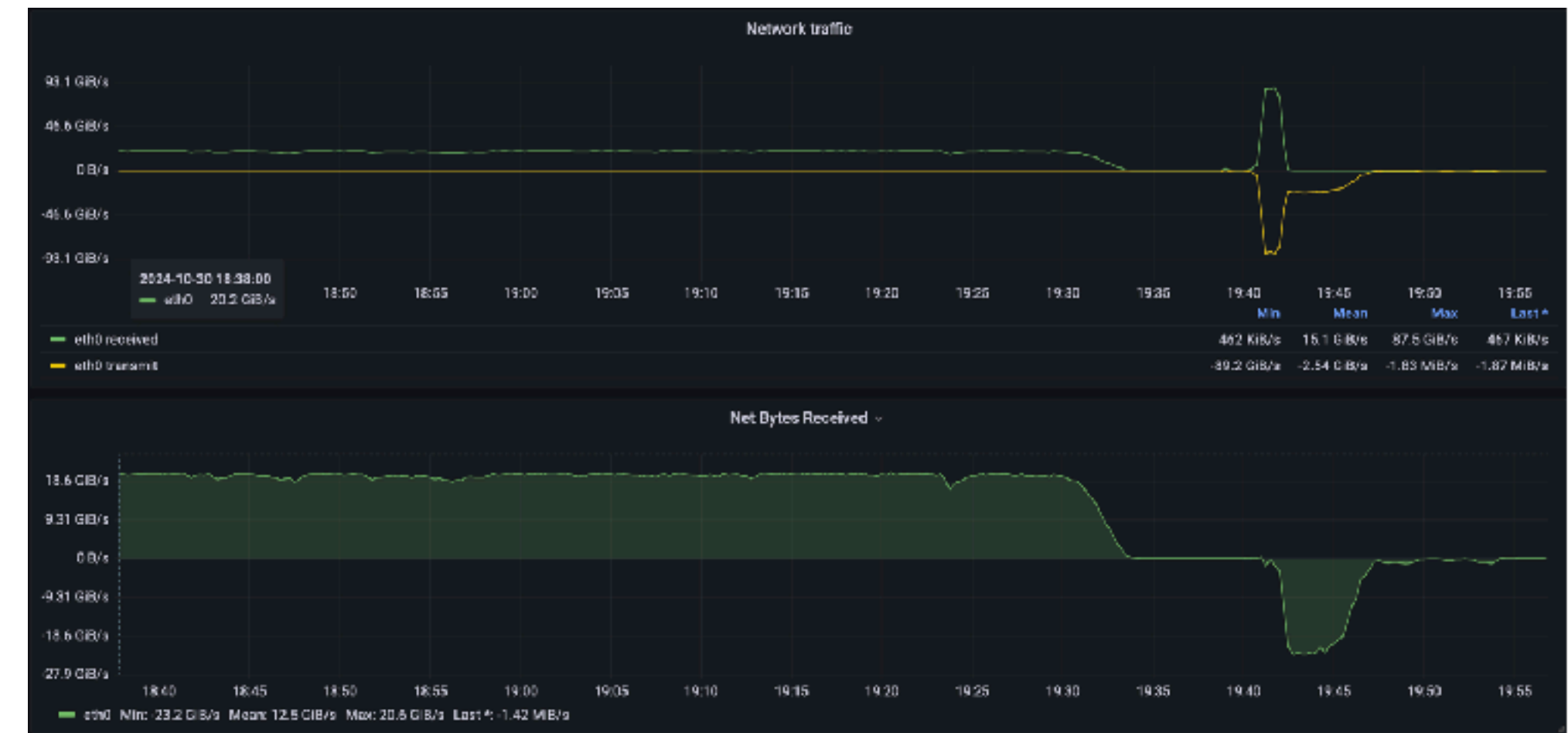
- Initial transfers from Arcus to Madrid had highly variable performance and were averaging 400-500Mb/s.
- Resolved through modifications to University firewall rules and use of SR-IOV port for external links to avoid virtual network overheads.

Data storage

- 2941TB Ceph storage on Arcus cloud (Cambridge) 2024-2025 allocation.
 - 1200TB triple replicated Ceph.
 - 1741TB 8+3 erasure coded Ceph.
- 1080TB S3 at RAL (Echo).
- Ceph storage is 80-90% full due to the recent completion of processing for DR4.
- Echo S3 storage is currently underutilised and only 21% full. Due to extended period of DR4 processing. Transfers of data to cold storage will be taking place over next few months and we expect this to increase dramatically by Feb 25.

Ceph

- From 2941TB total Ceph allocation, <1% is Cinder block device.
- 99% allocated in Openstack as Shares (Manila).
- Distributed processing requires multiple instance access to data for read. Spark designed to write to shared filesystems.
- Easy to set up administratively enforced read-only access to data.
- Total achievable bandwidth to/from Ceph is approximately 20GiB/s for cluster processing.
- However, this can impact other users of the cloud. And impact our own processes,



Cybersecurity

- Major push to improve cyber security underway in the University.
- New Cyber security policy published in April 2024 with total compliance required within 2 years.
- Many steps were already taken e.g. separation of system logs, ssh restrictions on source IP and specific users, firewall on all systems. Our systems have always been on separate private networks with limited specific servers providing external connections.
- Policy introduces many new requirements, especially around processes, audits, reviews etc. which will require extra effort.

Migration for 24/25

- Gaia Core Processing is starting to migrate to new systems on Arcus as part of moves to provide 24/25 allocations to all IRIS tenants on Arcus.
- Sapphire Rapids: Dual Intel Xeon Platinum 8468, total 184 vCPU, 992GiB RAM available for instances. 3.5TiB local storage. 100GbE.
- Memory per vCPU moves from 3.2 to 5.5GiB.
- End state is likely to be a mixture of 32 new systems and 35 of the current Icelake systems.
- This will require minor developments in our processing systems to efficiently utilise the mixture of hardware.

2025 and beyond (1)

- Finish modifications to migrated deployment. Benchmark and tune processes to improve utilisation of cloud resources.
- Finally migrate from Standalone Spark execution to Kubernetes to orchestrate execution. This has been postponed due to the extended period of processing for DR4.
- Software development and testing, including large scale testing, of new software in preparation for DR5 processing. Software version upgrades e.g. from Java 11 to Java 17.
- Experimental executions of proposed new processing to resolve issues/improve on DR4 results.

2025 and beyond (2)

- Begin processing for DR5. Schedule is still TBC. An estimated schedule used internally for planning:
 - Start major processing during 2026. Continuing through 2027 and final catalogue product generation for PhotPipe finishing during 2028.
 - Execution of SEAPipe processing during 2028 and 2029.
 - Post operations analysis and processing during 2029 and 2030.
- However, this is not a confirmed schedule and is expected to change over the coming months.