

Somerville: a high performance data intensive storage service for astronomy

The Somerville on-premises cloud platform is designed to support data-intensive use cases from survey astronomy, including very large databases (with trillions of records) and real-time data streaming and analysis.

Enabling unprecedented research

Named in honour of Scottish mathematician and astronomer Mary Somerville, the Somerville service will support thousands of astronomers internationally to pursue their science ambitions.

Since its initial installation in 2017, as part of the UK Science and Technology Facilities Council (STFC) e-Infrastructure pilot programme (now called the IRIS programme), Somerville has hosted production science services for the Vera C. Rubin Observatory and the Wide-field Astronomy Unit. It is also used as a testbed by the Gaia satellite, which is completing a census of more than one billion stars in the Milky Way, and by STFC's Research Cloud team.

At the time of writing, the largest database hosted on Somerville has more than 120 billion records of astronomy observations – that is 15 records for each and every person living on Earth.

Vera C. Rubin Observatory

The Rubin Observatory will conduct a ten-year survey of the Southern Hemisphere sky (the Legacy Survey of Space and Time, or LSST) using the world's largest digital camera. Rubin's science goals include understanding the nature of dark matter and dark energy, creating an inventory of the Solar System, mapping the Milky Way, and exploring objects that change position or brightness over time.



The 3200 megapixel camera that will take images for Rubin Observatory's 10-year Legacy Survey of Space and Time. (Jacqueline Ramseyer Orrell/SLAC National Accelerator Laboratory).

Facilitating data science

The volume and complexity of data that Rubin will capture will challenge the capabilities of research computing worldwide. Somerville will be the host for the UK's Rubin Independent Data Access Centre (DAC) and will connect astronomers with at least 200 Petabytes of survey products and provide advanced tooling to handle these products.

EPCC, in collaboration with the Institute for Astronomy at the University of Edinburgh, has a critical role in the design and operation of the UK DAC, which is due to enter operation in 2025.

Technology

Somerville is installed at the University of Edinburgh's Advanced Computing Facility, and managed by a joint team from EPCC and the University's Institute of Astronomy.

The system is built on Scientific OpenStack. This variant of the popular OpenStack toolkit has been customised by StackHPC Limited to meet research-computing capabilities.

Nodes

- 4,224 (virtualised) CPU cores.

Interconnect technologies

- 100G Ethernet data network
- 2 x 100G uplink into JANET.

System storage

- 5PB of Ceph storage for holding scientific data
- 100TB of NVMe-based Ceph storage.

Funding

Somerville is funded by STFC via the LSST:UK and IRIS programmes.



Somerville at EPCC's Advanced Computing Facility.

Rubin Independent Data Access Centre

The Rubin Independent Data Access Centre (DAC) provides three different user interfaces to LSST:

- A web-based science portal for interactive enquiries
- A notebook-based engine for more substantial and scripted analysis workflows
- An HPC/batch interface for the most ambitious and computationally intensive research campaigns.

These interfaces are backed by custom services, engineered within the Rubin project to the requirements of the survey. They include a bespoke distributed database system called Qserv, able to serve LSST catalogues containing hundreds of billions of objects and to satisfy the demands of hundreds of concurrent queries.

Further information

Rubin Observatory

<https://rubinobservatory.org>

Wide Field Astronomy Unit

<https://www.roe.ac.uk/ifa/wfau/>

Gaia

https://www.esa.int/Science_Exploration/Space_Science/Gaia

LSST:UK

<https://www.lsst.ac.uk>

IRIS

<https://www.iris.ac.uk>

Access

To access Somerville, please see:
<https://www.iris.ac.uk/rsap>