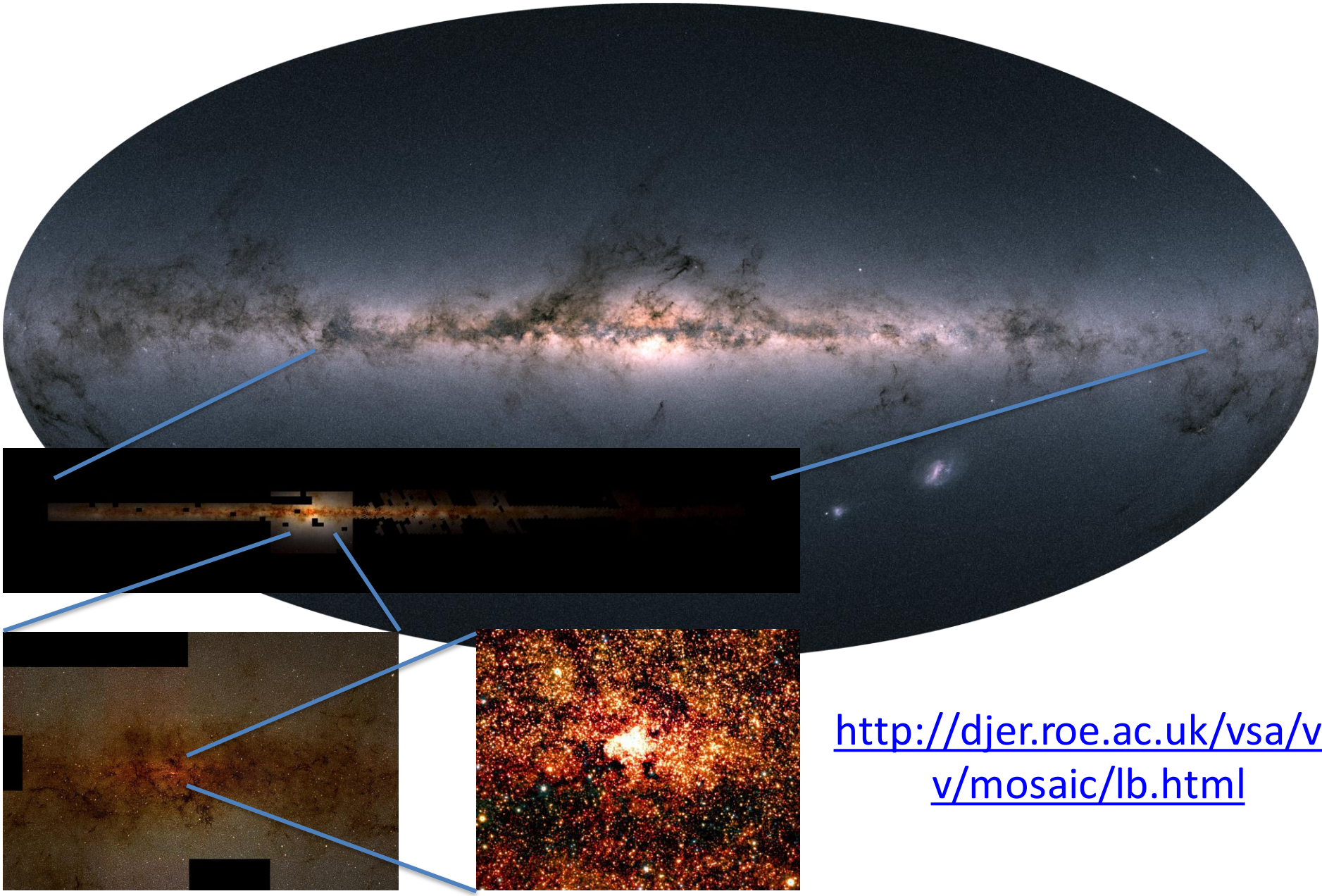
Space is big. Really big. So is the data.

(… you just won't believe how vastly, hugely, mind-bogglingly big …)

Nigel Hambly
Wide Field Astronomy Unit, Institute for Astronomy
School of Physics and Astronomy
University of Edinburgh

Big Surveys = Big Data



http://djer.roe.ac.uk/vsa/vvv/mosaic/lb.html

WFAU at IfA

- Formed in 1999, built on the long tradition of astronomical imaging science and research at the ROE on Blackford Hill
- Production and digitisation of legacy photographic surveys
- Progressed to managing data from new generation optical / near infrared surveys
- Latterly involved in data processing SW engineering for large space-based survey missions (European Space Agency)
- Now preparing for the next generation of ground and space-based surveys
- Recently moved into new offices in the Higgs Centre for Innovation, co-located with the UK Astronomy Technology Centre
- Current head count: 14 Research Software Engineers

WFAU projects: past

- SuperCOSMOS digitisation/dissemination programme (pre-2008)
  - http://ssa.roe.ac.uk
  - 20 TB imaging data
  - 4 TB, 10^{10} row catalogues (all sky)

- United Kingdom Infrared Sky Surveys science archive
  - http://wsa.roe.ac.uk
  - 230 TB imaging data
  - 17 TB, 10^{10} row catalogues (¼ sky)

- Astrogrid: SW infrastructure for the global Virtual Observatory
  - http://www.astrogrid.org

WFAU projects: present

- United Kingdom Infra-Red Telescope Hemisphere Survey
- European Southern Observatory's visible and infrared survey telescopes public surveys
  - http://osa.roe.ac.uk
  - http://vsa.roe.ac.uk
  - 310 TB imaging data
  - 136 TB, 10^{11} row catalogues (½ sky)
- Gaia-ESO spectroscopic surveys
  - http://ges.roe.ac.uk
- Gaia (European Space Agency)
  - http://sci.esa.int/gaia/
  - ≈ 100 TB data
  - 10^{12} row catalogues (all sky)
- Further SW infrastructure work for the Euro-VO and International Virtual Observatory Alliance
  - Tools, infrastructure and standards for inter-operability

WFAU projects: in prep. / future

- ESA-Euclid
  - http://sci.esa.int/euclid/
  - ≥ 1000 TB imaging data (≥ ½ sky)
  - systems engineering, data processing SW engineering, Data Processing Centre preparation

- Large Synoptic Survey Telescope
  - https://www.lsst.ac.uk
  - ≈ 60,000 TB imaging data (≥ ½ sky)
  - UK Data Access Centre preparations

The IT Future for WFAU …

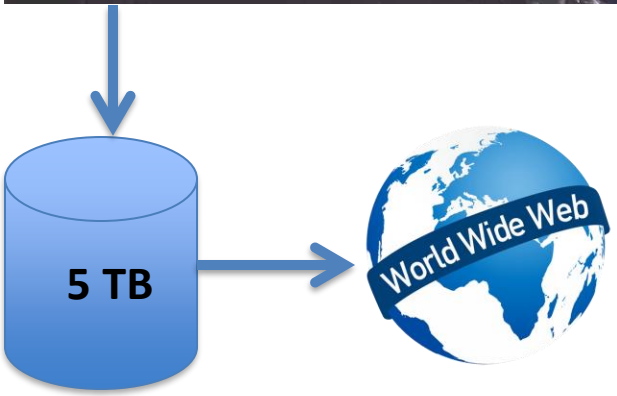
- Larger collaborations
  - International survey projects
  - Professional Research Software Engineering
- Data management
  - No longer purely in-house
  - Distributed systems
- Computing model is evolving
  - IT base no longer purely relational
  - Scale-out usage scenarios need HPC / HTC
  - Cloud and Grid-based technologies

The importance of RSEs

- Data processing
  - Research background = domain knowledge
  - IT training = professional programming standards
  - Collaborative development
- Data serving
  - Familiarity with state-of-the art technologies
  - Abilities in system administration and network infrastructure
  - Maintenance of high availability services
- Data preservation
  - Long term curation, e.g. migrate data onto new media
  - Ensures survey legacy value is maintained

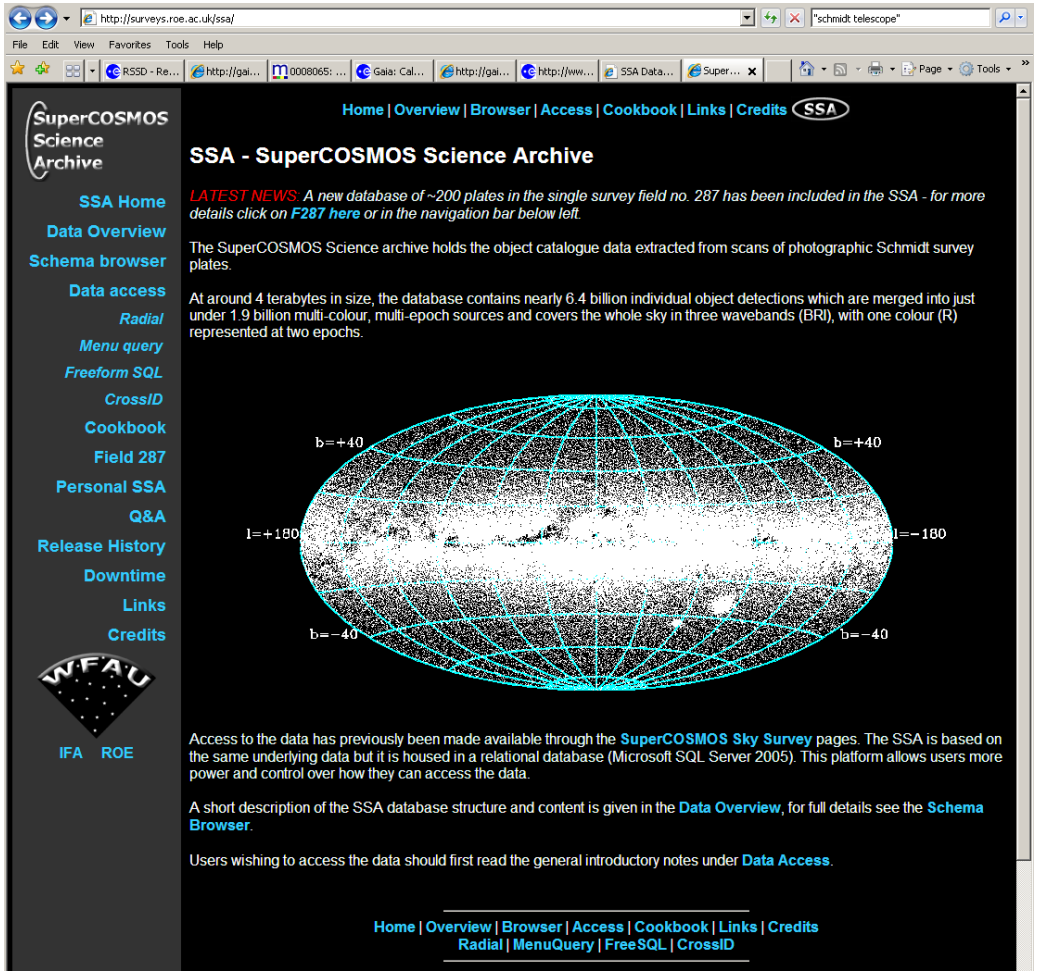
SuperCOSMOS Science Archive

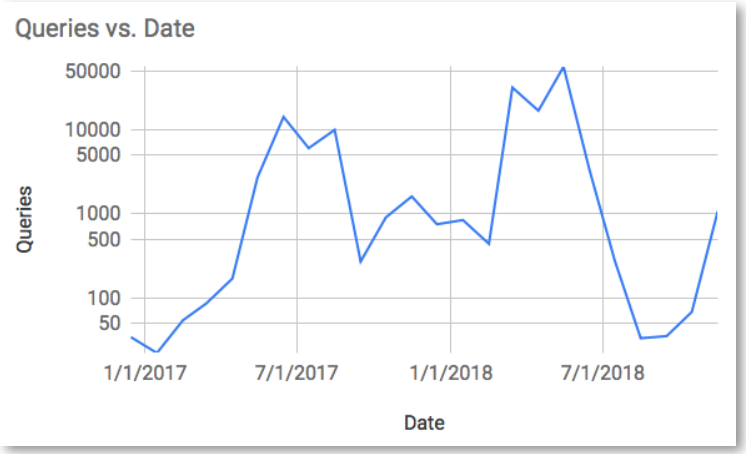
Data preservation (I)

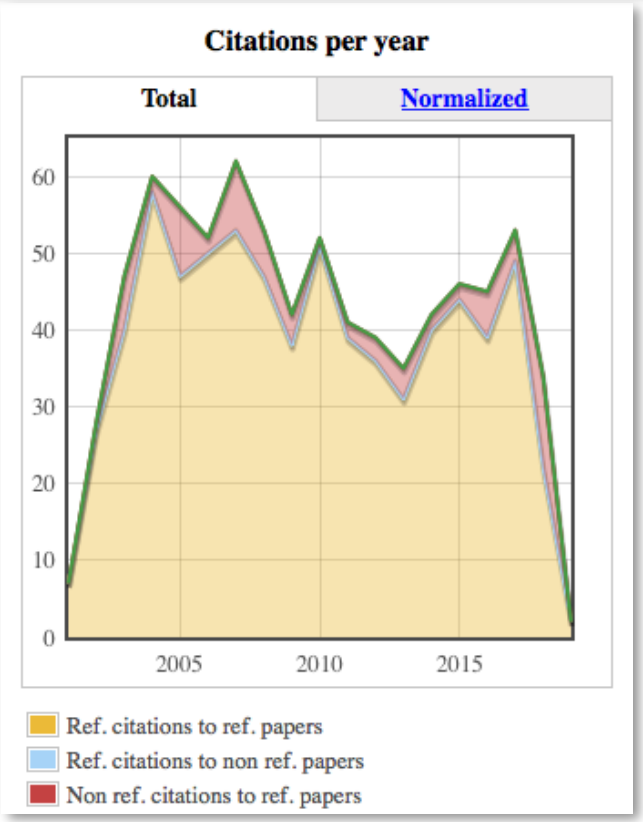


Data preservation (II)

http://ssa.roe.ac.uk







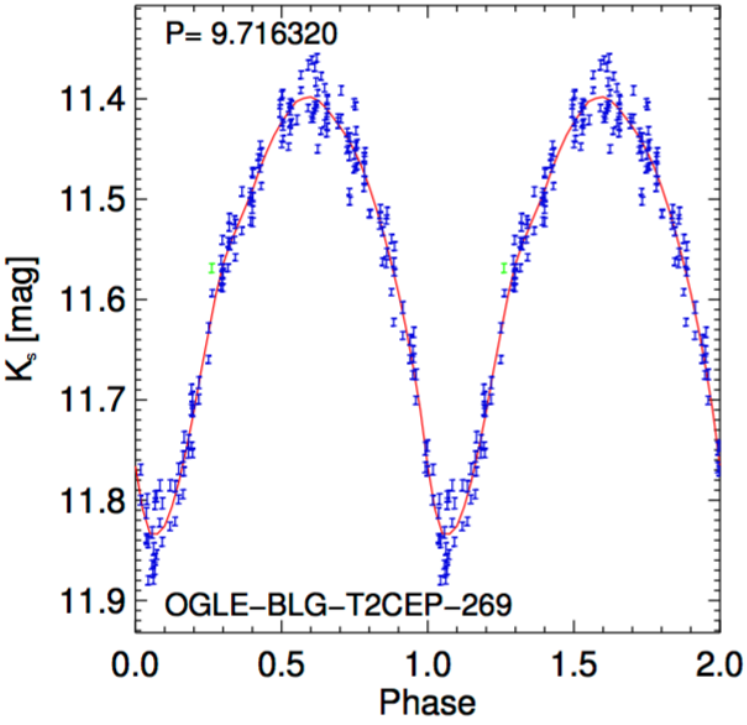
(… plus 1000s of image requests per week …)

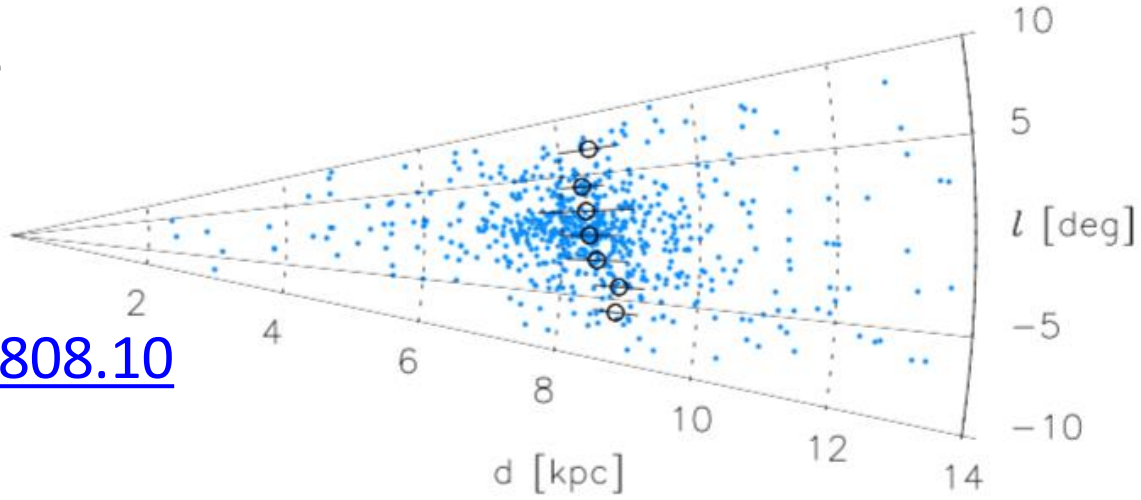
Data serving (I)

- e.g. VISTA Variables in the Via Lactea (VVV)
  - Time domain infrared imaging survey
  - Yields light curves, stellar motions, …
- $\approx 10^9$ unique sources with $\approx 10^{11}$ epoch measures (5^{th} Data Release)
  - 60 billion row match table
  - Final DR “VVVX” 2x bigger
- http://surveys.roe.ac.uk/vsa

Data serving (II)

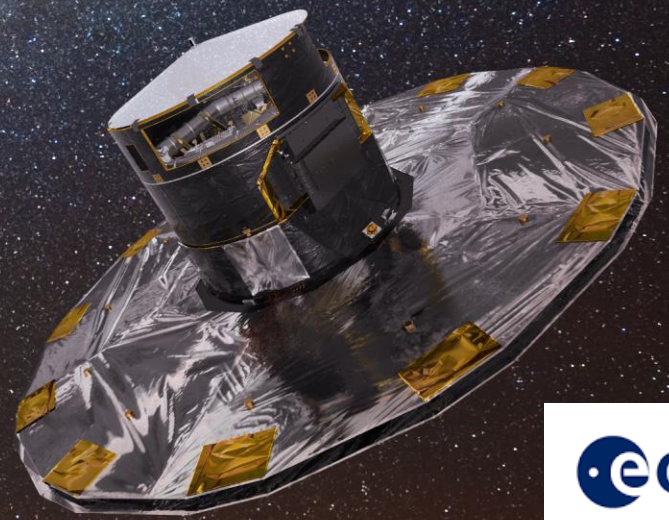
- Recent science highlight: Braga et al. (2018)
  - Infrared imaging penetrates the interstellar gas and dust (which obscures visible light)
  - Time-resolved measurements yield light curves
  - Identify “Cepheid” variables which have a Period-Luminosity correlation
  - Spatial structure of the Galactic Bulge mapped
  - https://arxiv.org/pdf/1808.10838.pdf





Data processing (I)

- European Space Agency's Gaia mission
  - Launched December 2013
  - WFAU staff: 3 (out of 100s spread around Europe)
  - Provide core processing SW for the "ground segment"
  - Data rate: up to 8 Mbit/s
  - Key goal: 3-d view of the Galaxy via trigonometric parallax measurements for 10^9 stars

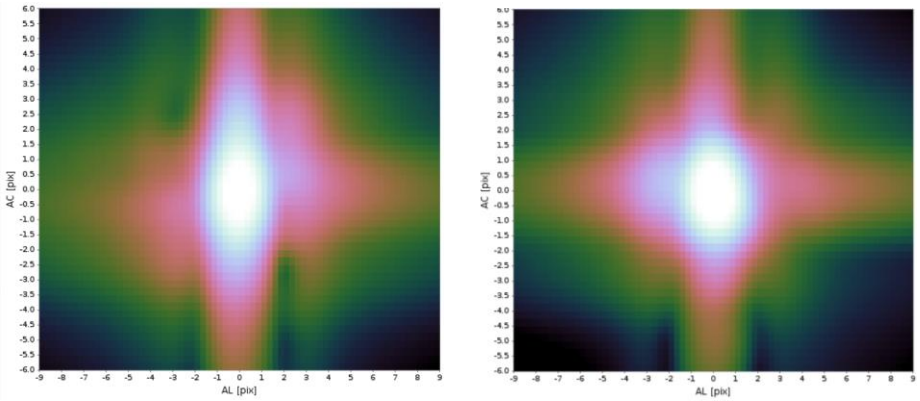


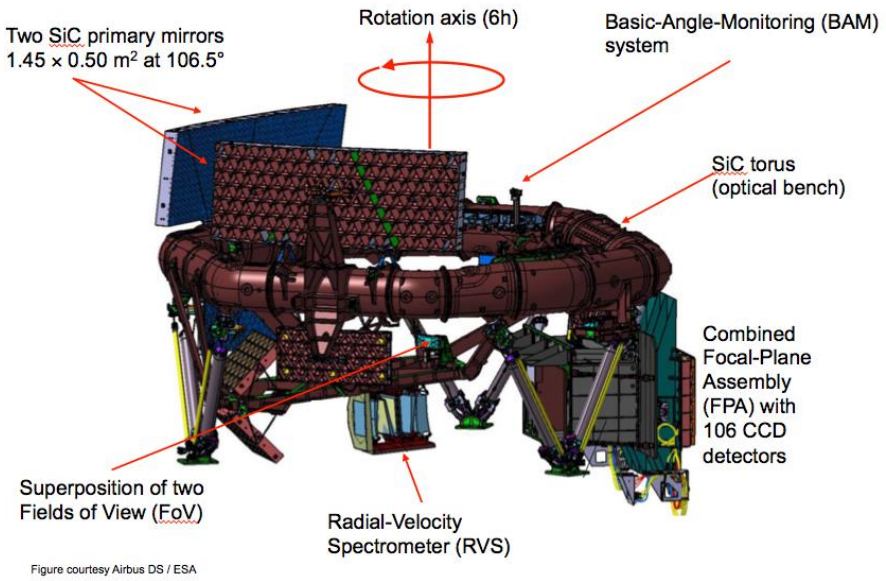
GAIA MISSION STATUS

1592 days in routine phase
60,065 GB of science data gathered
115,062,237,531 transits observed

esa
European Space Agency
Agence spatiale européenne

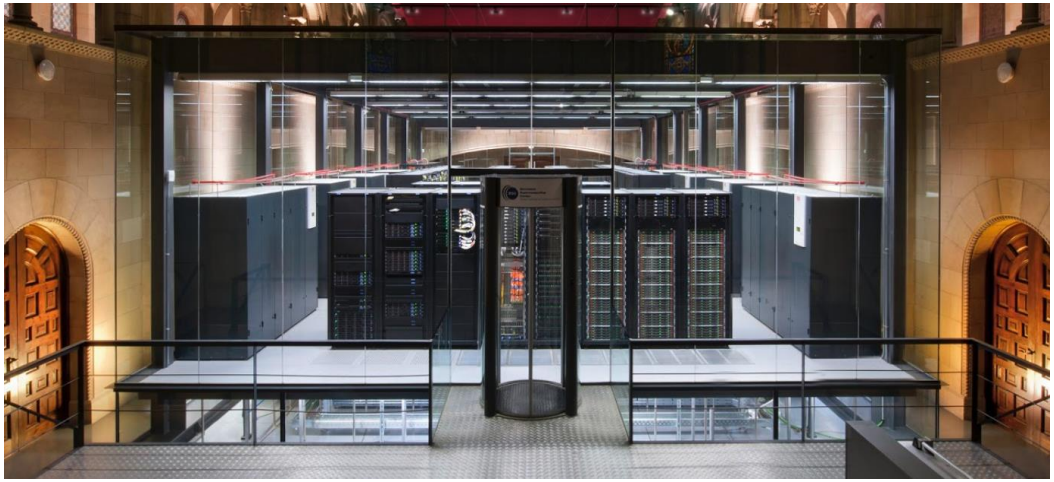
Data processing (II)



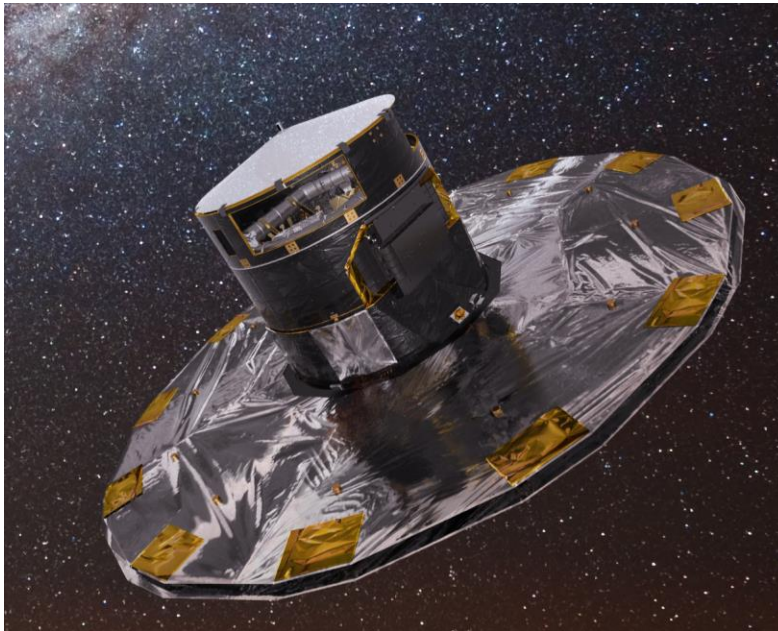


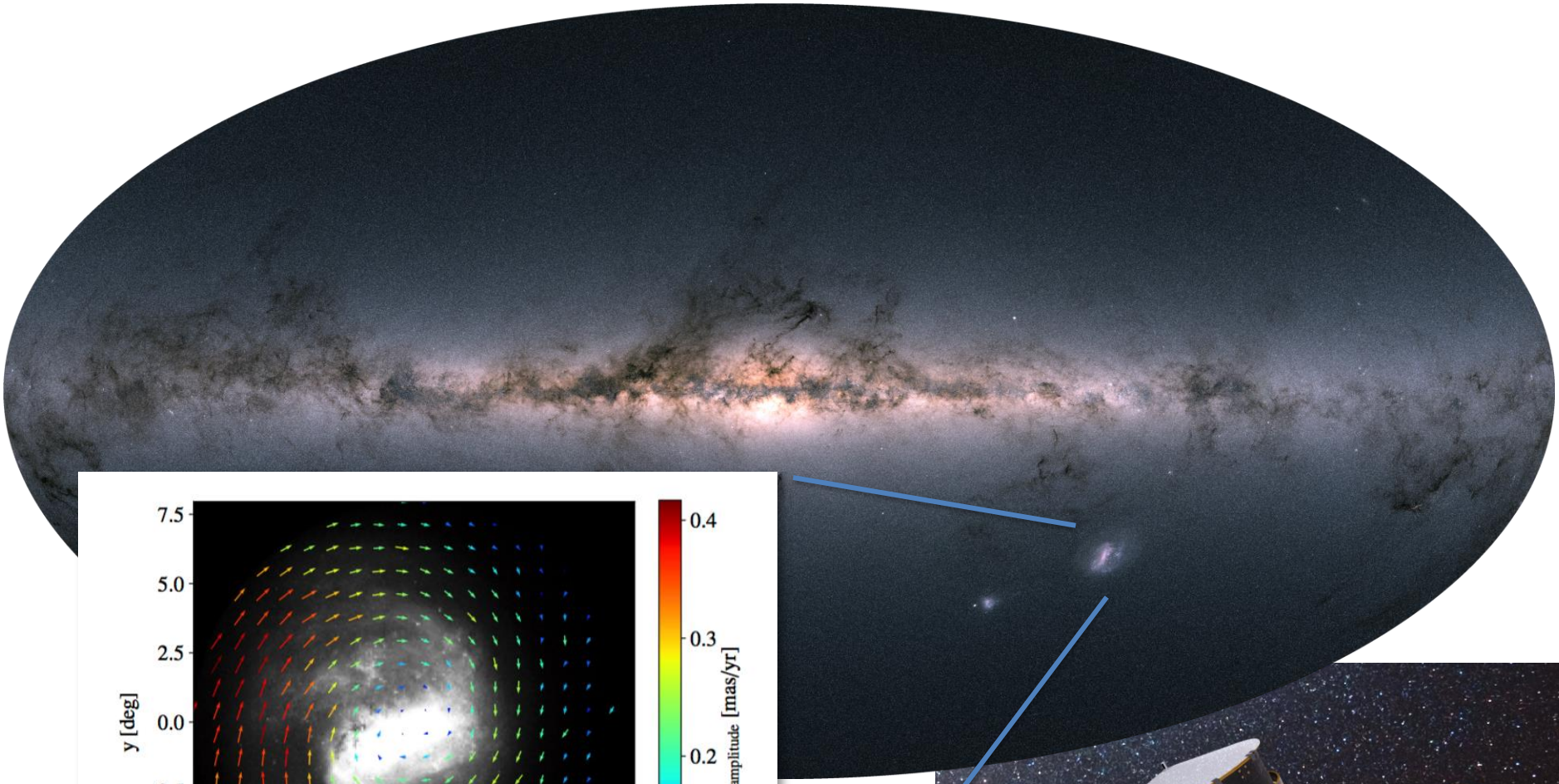
WFAU staff provide key instrumental calibration software for Gaia:

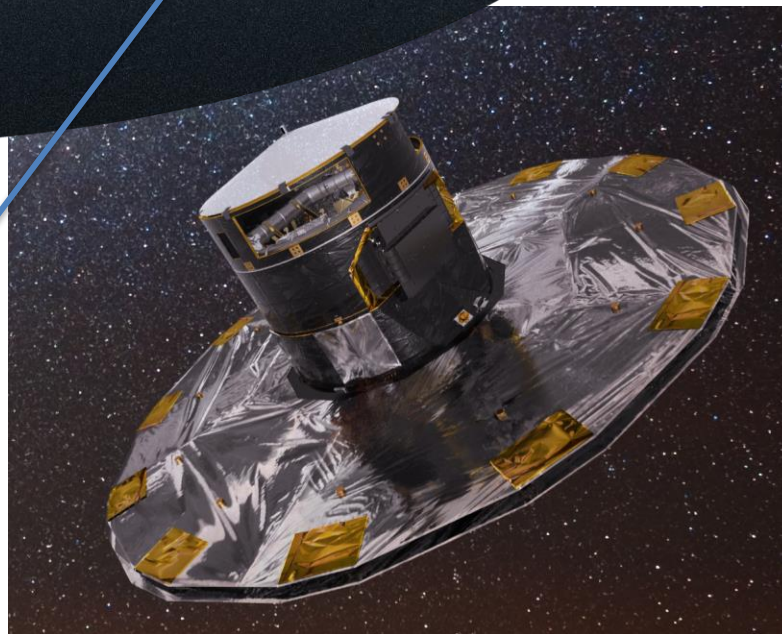
- Optical
- Electronic



WFAU SW runs at processing centers around Europe

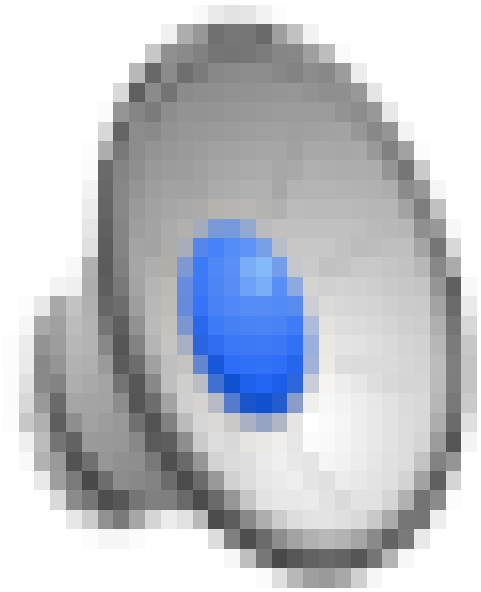






Helmi et al. (2018)
https://arxiv.org/pdf/1804.09381.pdf

Our locality in three dimensions ...



Acknowledgement: Gaia Data Processing and Analysis Consortium (DPAC); Gaia Sky;
S. Jordan / T. Sagristà, Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Germany


Merry
Christmas