SPAN

SPACE ACADEMIC NETWORK

Nicholas Walton

(Institute of Astronomy, University of Cambridge)

SPAN Data and Analysis Working Group

White Paper on Space Data

Excellence in Growth: Space Data Infrastructure for a Competitive UK

SPAN SPAN Data White Paper:

Presentation slides

The SPAN white paper 'White Paper on Space Data Excellence in Growth: Space Data Infrastructure for a Competitive UK' is available for download from
https://span.ac.uk/publications/
at:
https://span.ac.uk/wp-content/uploads/2026/03/span-data-whitepaper-v1.0.pdf

SPAN Data and Analysis Working Group

- This paper developed by the SPAN SDAWG
  - Brings together academics working across space science in the widest sense, with an interest or background in 'data'
    - University academics involved in the development and operations of space science and earth observation missions, especially in the development of the data processing
    - Academics and software engineers from National Laboratories (e.g. RALSpace, UKATC)
    - Scientists and policy representatives from related groups, e.g. DSTL, MetOffice
    - Policy representatives from funding agencies, e.g. UKSA, STFC, NERC
- Paper includes input from the wider SPAN network
  - Working Groups: Science & Exploration, Engineering, Earth Observation

SPAN
SPACE ACADEMIC NETWORK

Why Space Data Matters

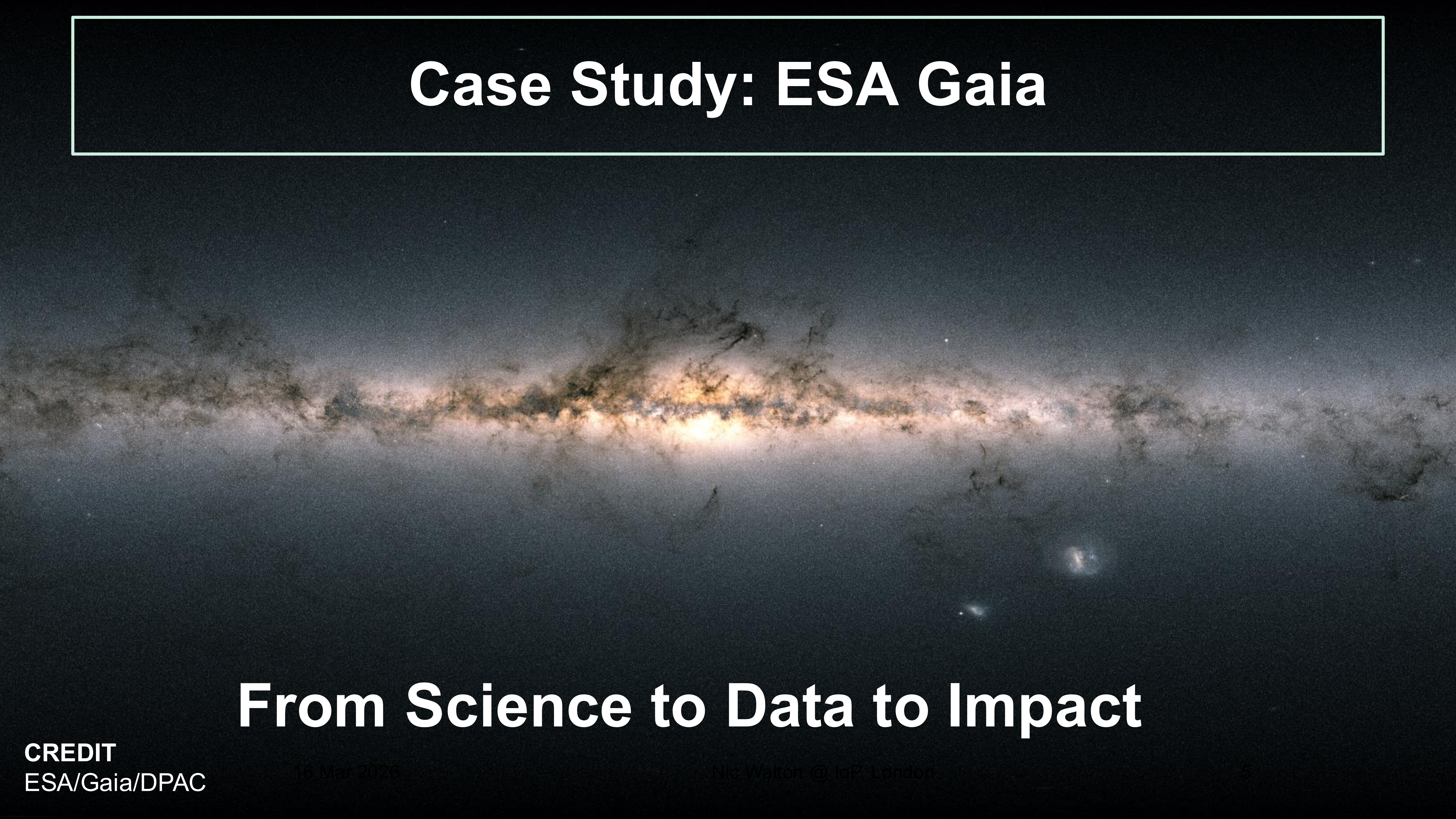
- Space is an increasing area of importance and is underpinning an increasing significant volume of economic activity
- The UK Space Agency currently organises its activity across 7 themes, with four now identified as higher priority

  - Launch and Space Transportation
  - Space Domain Awareness (SDA)
  - Satellite Communications
  - In Orbit Servicing, Assembly and Manufacturing (ISAM)
  - Space Science and Exploration
  - Earth Observation (EO)
  - Position, Navigation, and Timing (PNT)

- Within each of these activities Data infrastructure is the mechanism through which hardware investment generates economic return

Economic returns are significant: £1 invested returns £7.5

SPAN SPACE ACADEMIC NETWORK



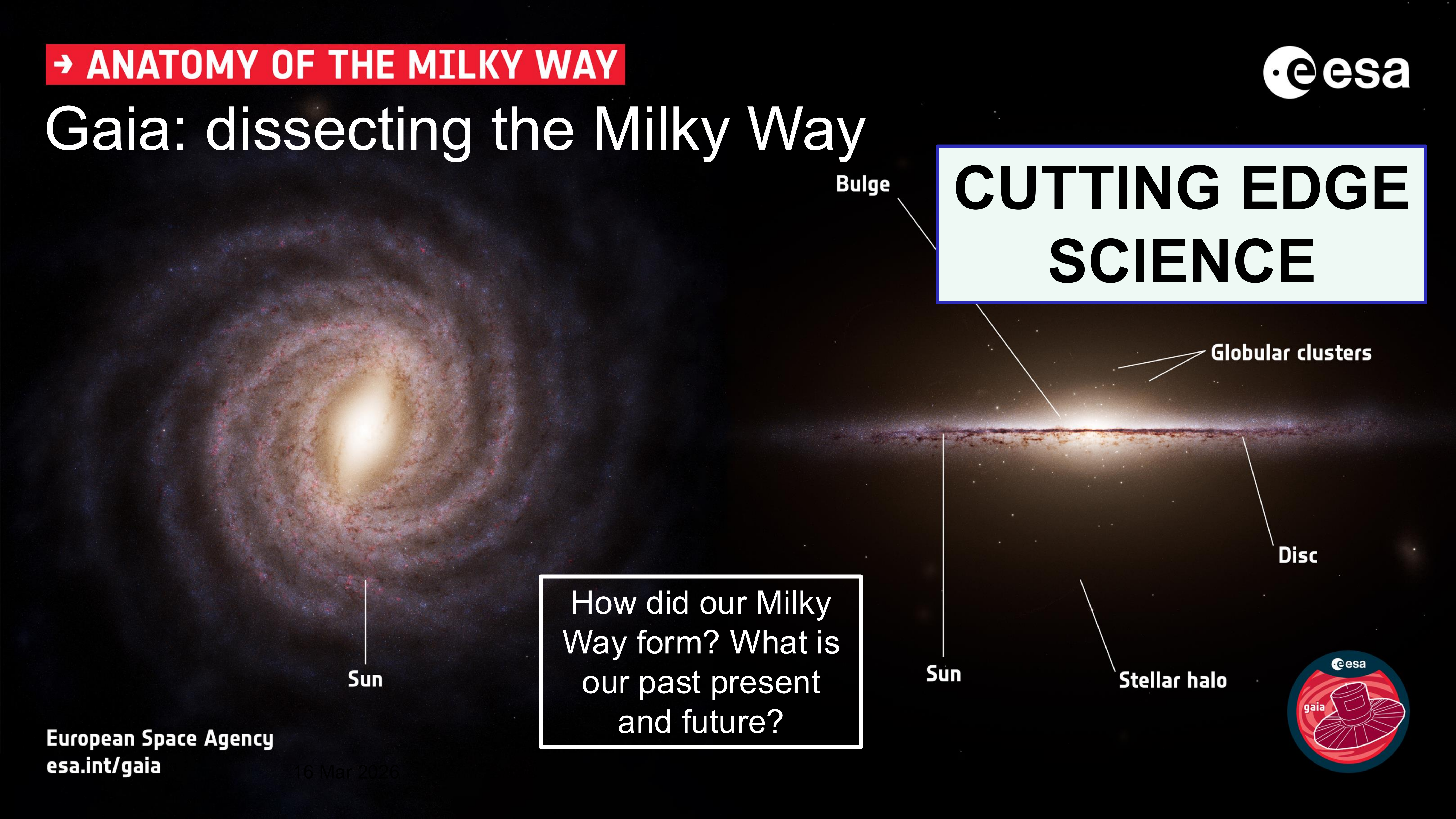
Case Study: ESA Gaia

From Science to Data to Impact

CREDIT
ESA/Gaia/DPAC

Gaia: dissecting the Milky Way

CUTTING EDGE SCIENCE



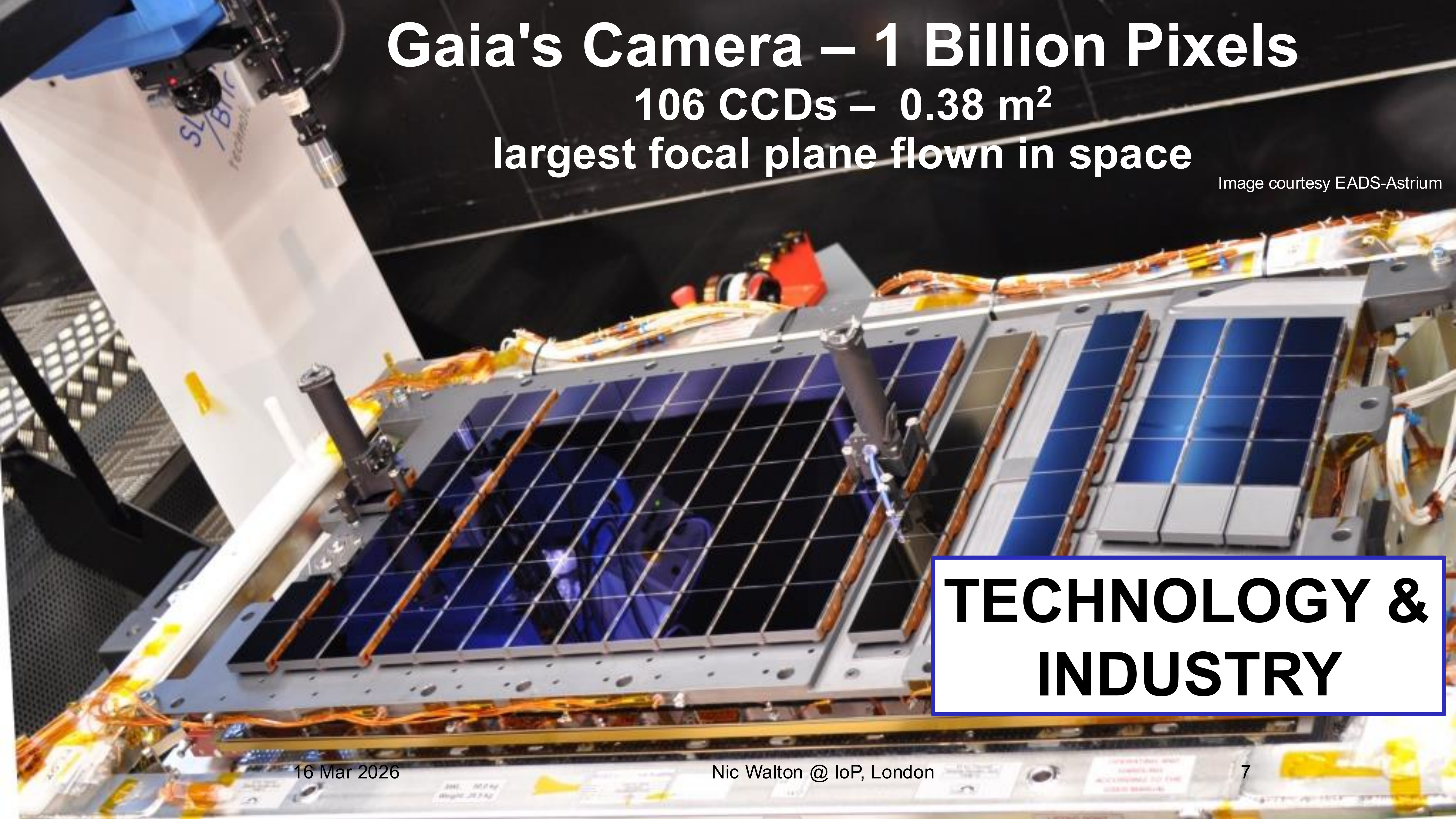
How did our Milky Way form? What is our past present and future?

Gaia's Camera – 1 Billion Pixels

106 CCDs – 0.38 m^2

largest focal plane flown in space

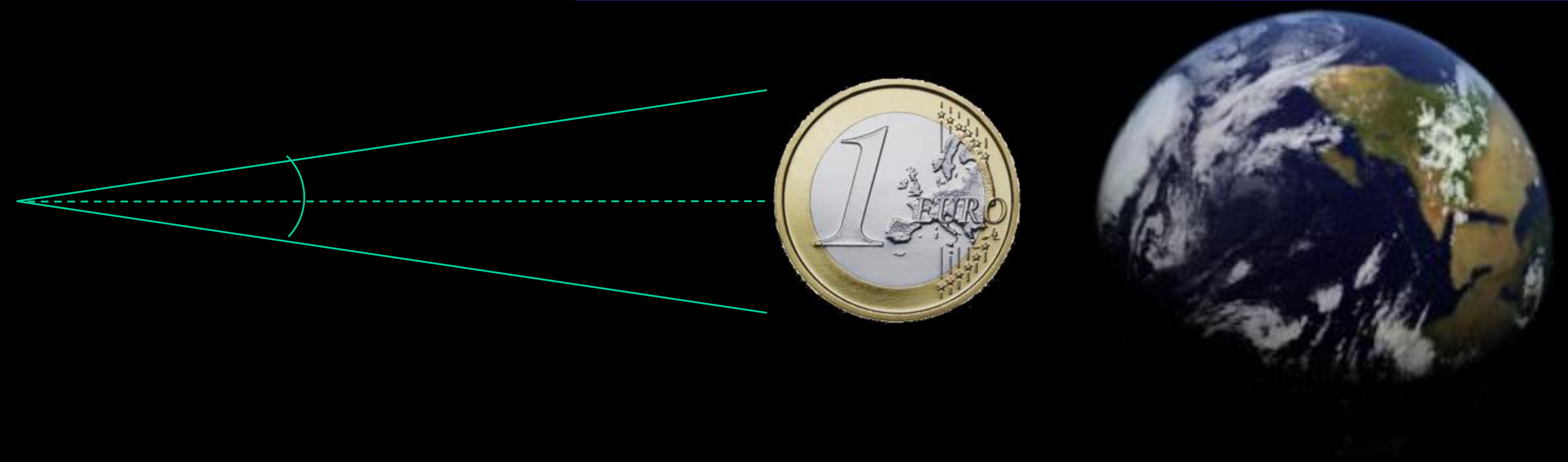
Image courtesy EADS-Astrium

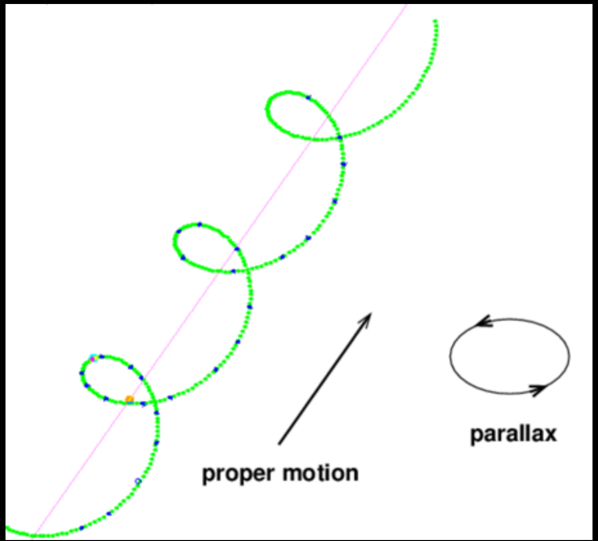


TECHNOLOGY & INDUSTRY

Gaia: Astrometry in the extreme

PUSHING THE ENVELOPE



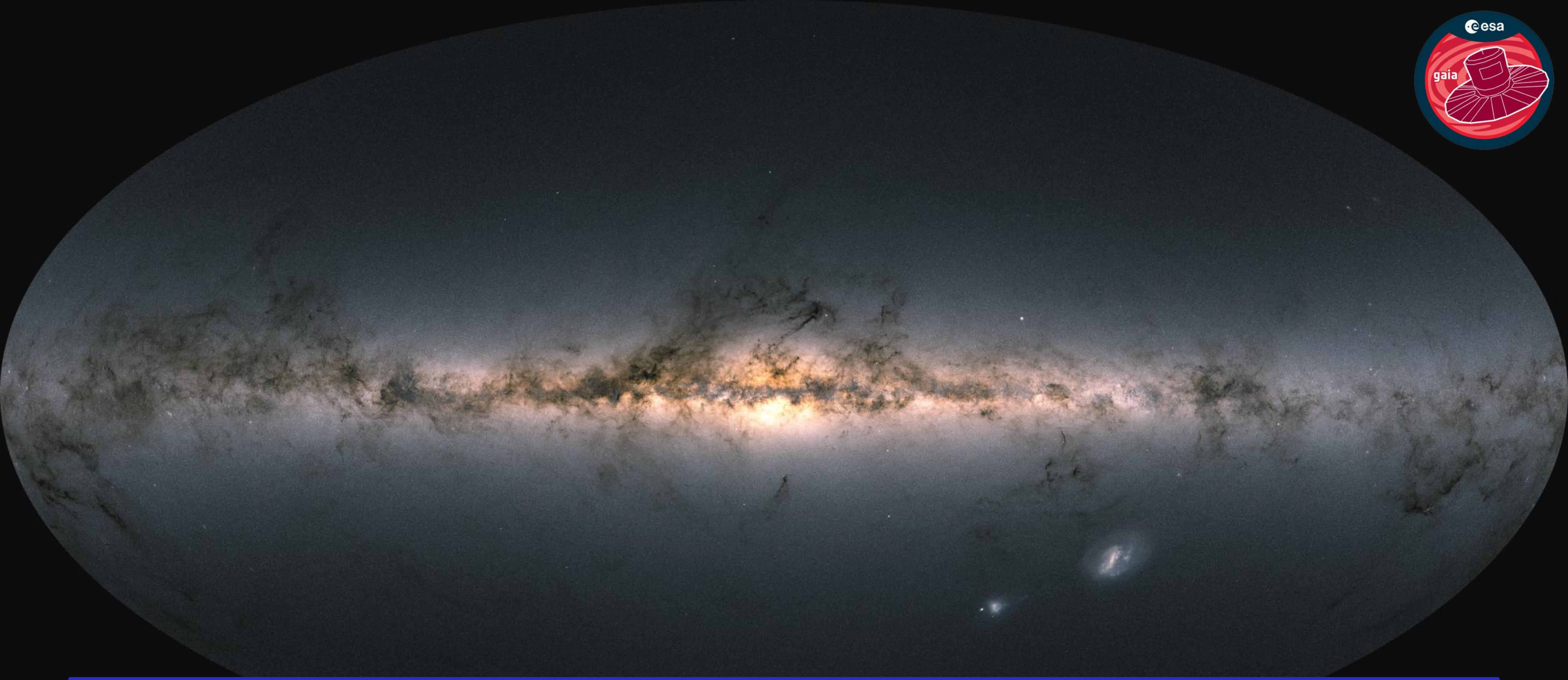


Gaia precision to ~10 micro-arcseconds

Enables distances to be measured with ~10% errors to ~10 Kpc (= ~ 32,000 light years)

This is the equivalent of a Euro coin on the earth as seen by the next human on the moon!

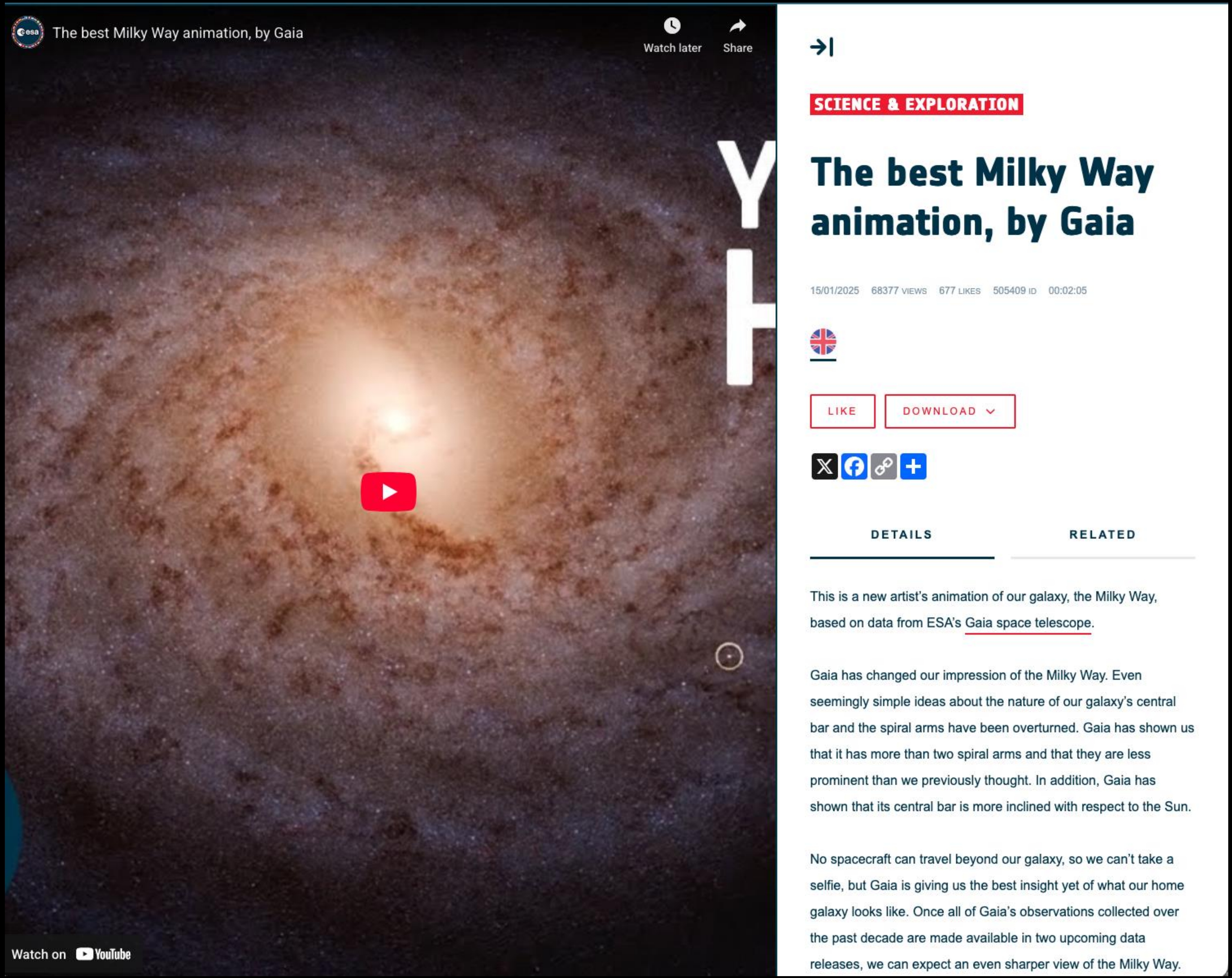
Moon/ Earth Credit: NASA



OPEN SCIENCE, AND MORE OPEN SCIENCE

Gaia EDR3 (~1.5 Billion) sources (Dec 2020) / Credit ESA/Gaia/DPAC

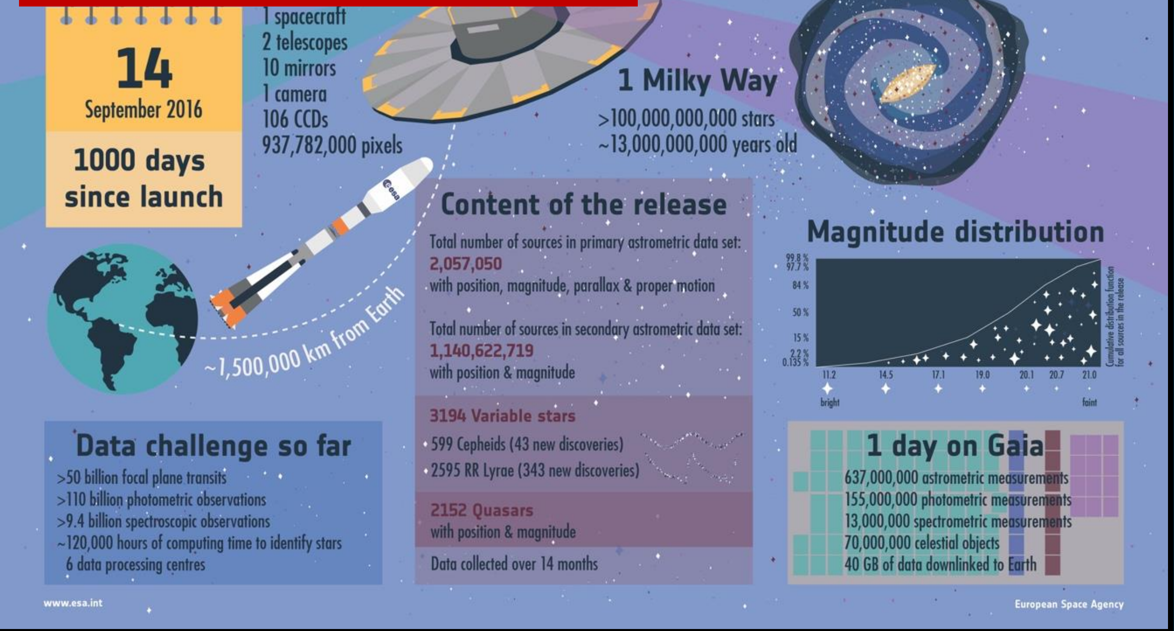
TRANSFORMATIONAL SCIENCE



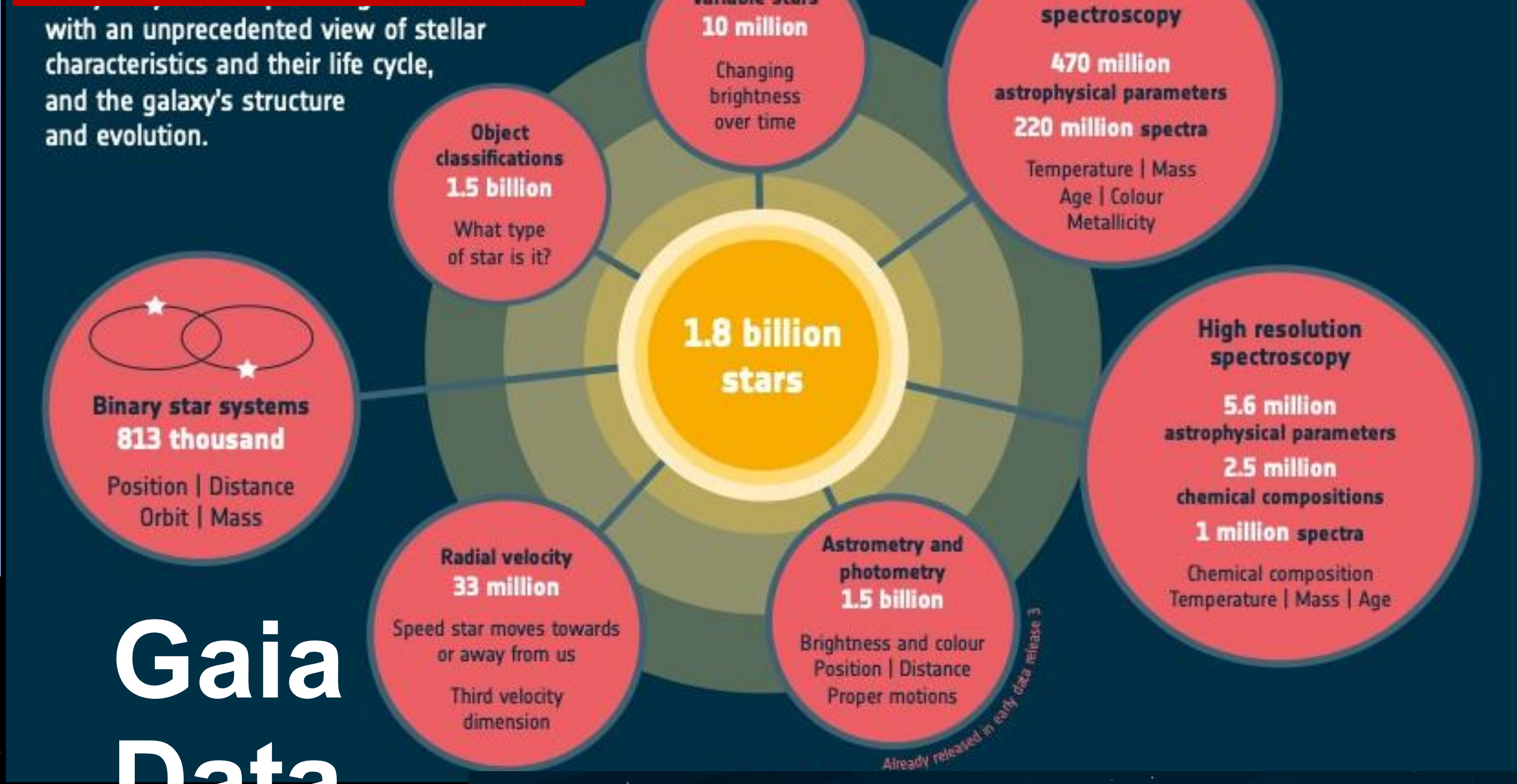
ESA GAIA VIDEO 'The best Milky Way animation, by Gaia' – can be download from https://www.esa.int/ESA_Multimedia/Videos/2025/01/The_best_Milky_Way_animation_by_Gaia

SPAN
SPACE ACADEMIC NETWORK

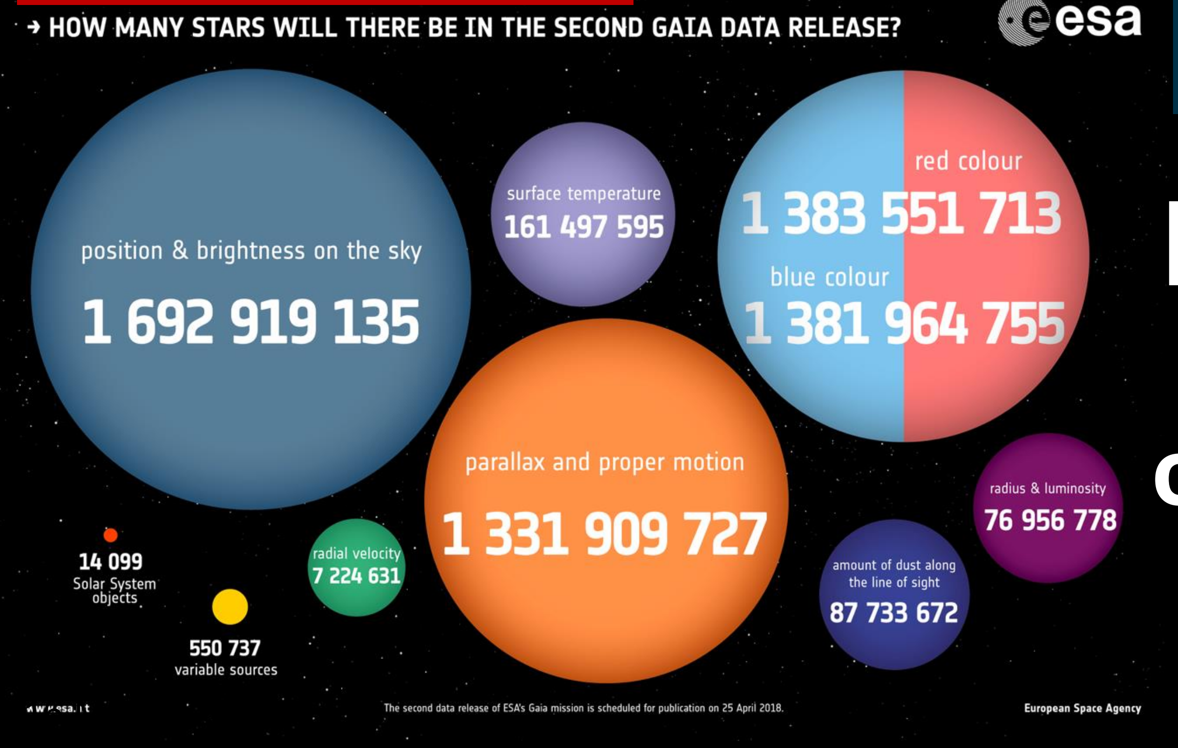
Gaia DR1:Sep 2016



Gaia DR3: Jun 2022



Gaia DR2: Apr 2018

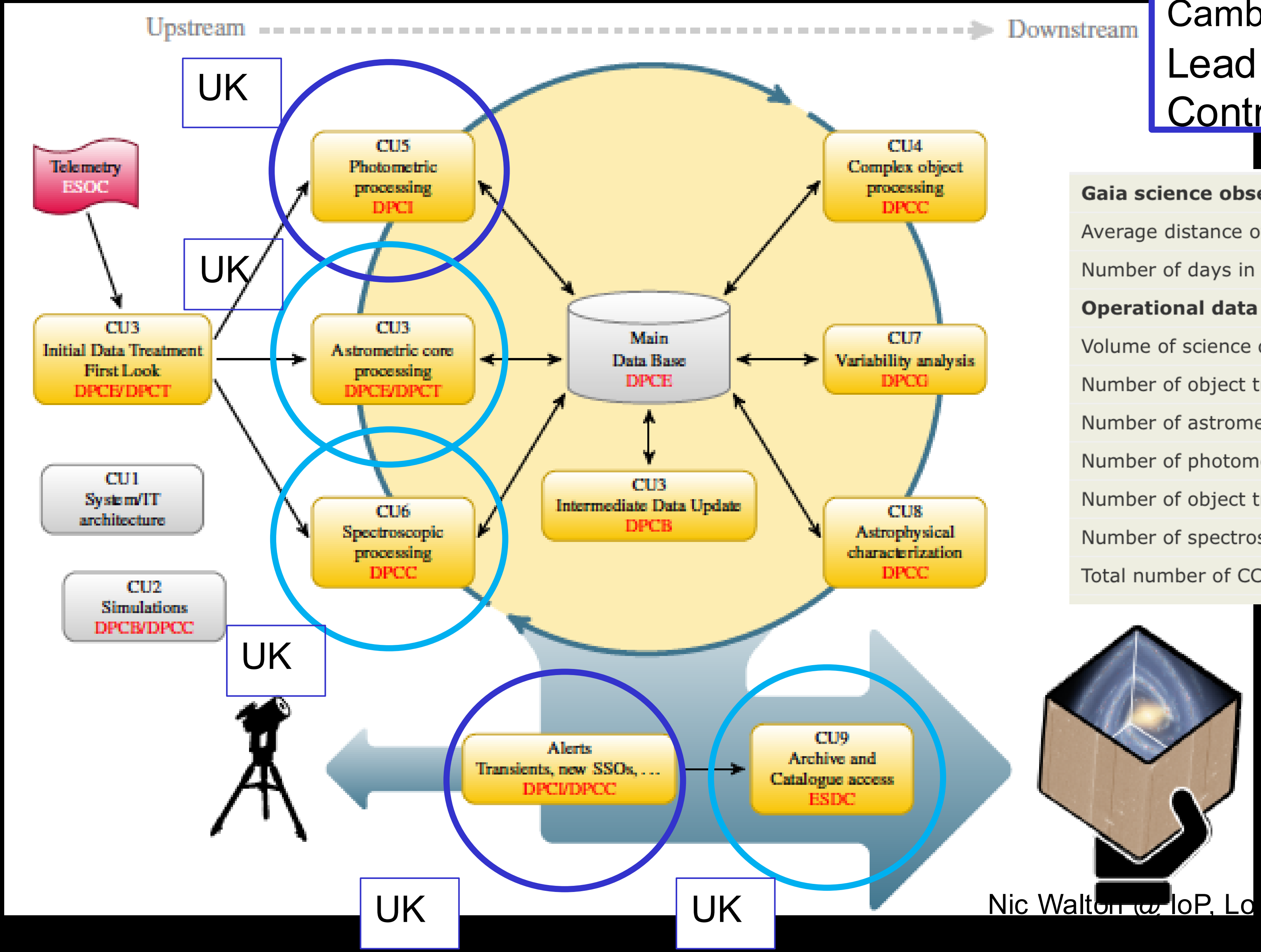


Gaia Data Releases increasing complexity, increasing size

Gaia EDR3: Dec 2020

Gaia Processing is complex

a pan European effort



A STRONG UK ROLE

Significant UK involvement:

Cambridge Gaia Data Processing Centre

Lead Photometric processing and Flux alerts

Contribute to pre-processing, RVS, archive

Gaia science observation phase	
Average distance of Gaia space telescope from Earth (in km)	1,510,000
Number of days in science operations (25 July 2014 - 15 January 2025)	3827
Operational data collected	
Volume of science data collected (in GB)	141,064
Number of object transits through the focal plane	267,356,167,925
Number of astrometric CCD measurements	2,635,367,940,969
Number of photometric CCD measurements	530,852,459,518
Number of object transits through the RVS instrument	17,476,697,079
Number of spectroscopic CCD measurements	52,066,956,051
Total number of CCD measurements (astrometric + photometric + spectroscopic)	3,218,287,356,538

3.2 Trillion observations made during the 3827 of inflight operations!

The People Behind Gaia Data Releases

PEOPLE MAKE IT HAPPEN

Now working on Gaia DR4 (Dec 2025, 66 months of data)
DR5 ~2030 (10+ yrs of data)

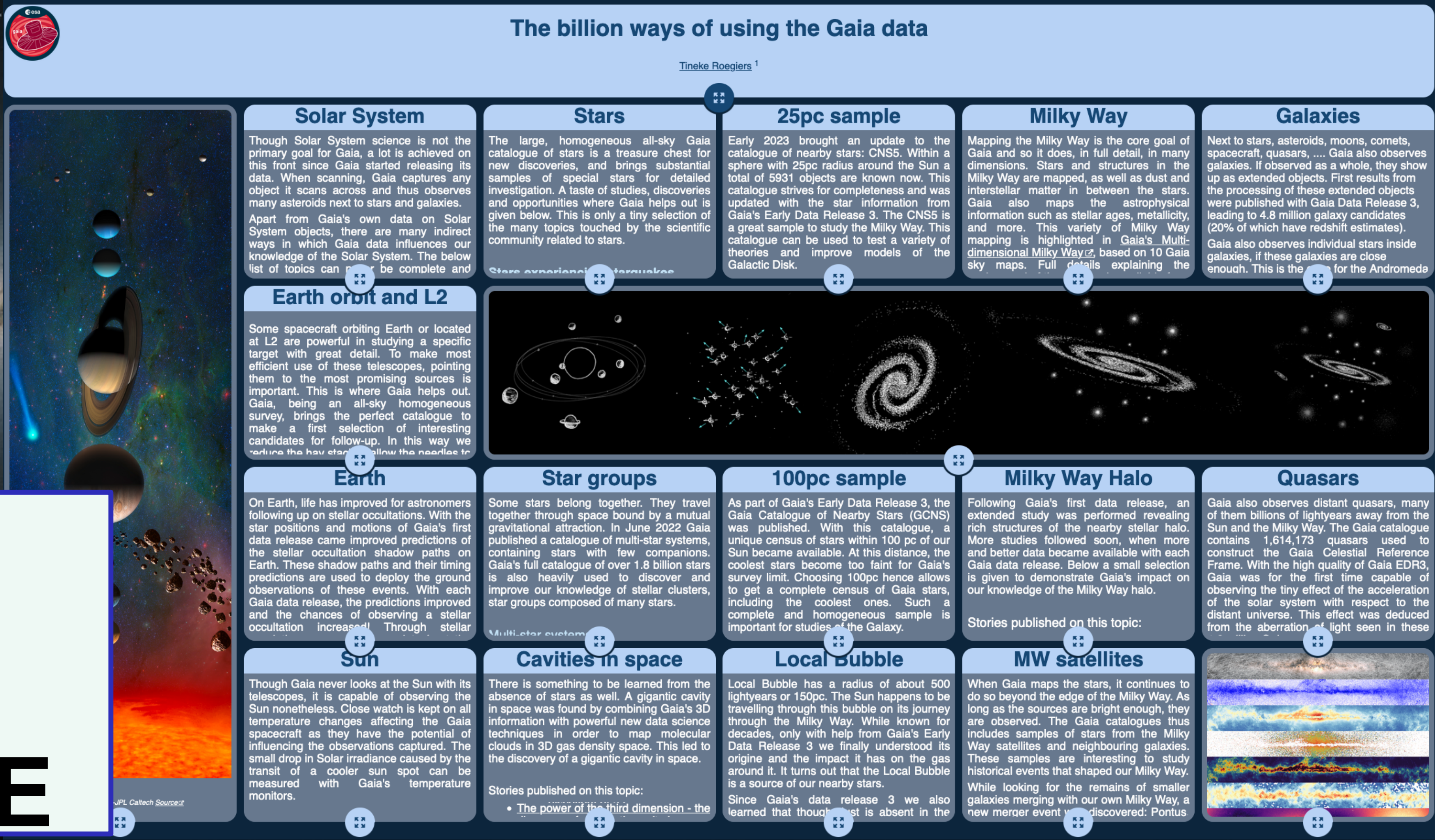
Gaia UK responsible for the processing and analysis of the photometric data, the spectrophotometric data, photometric science alerts and with essential contributions to the spectroscopic data

Gaia DR3 Science Examples

https://k-poster.kuoni-congress.info/eas-2023/poster/77628391-074f-4689-867f-79d6342116fc

Take a look at Tineke Roegiers' (Gaia ESA Outreach coordinator) EAS 2023 presentation: click through each link for more info, papers etc.

15,500 peer reviewed papers to date (Jan 2026) based on Gaia data



QUALITY DATA UNDERPINS QUALITY SCIENCE

ESA/Gaia/DPAC

Gaia DR4 and Beyond

https://www.cosmos.esa.int/web/gaia/release

66 months of data, ~2.5 Billion sources

Full astrometric, photometric, and radial-velocity catalogues.

All available variable-star and non-single-star solutions.

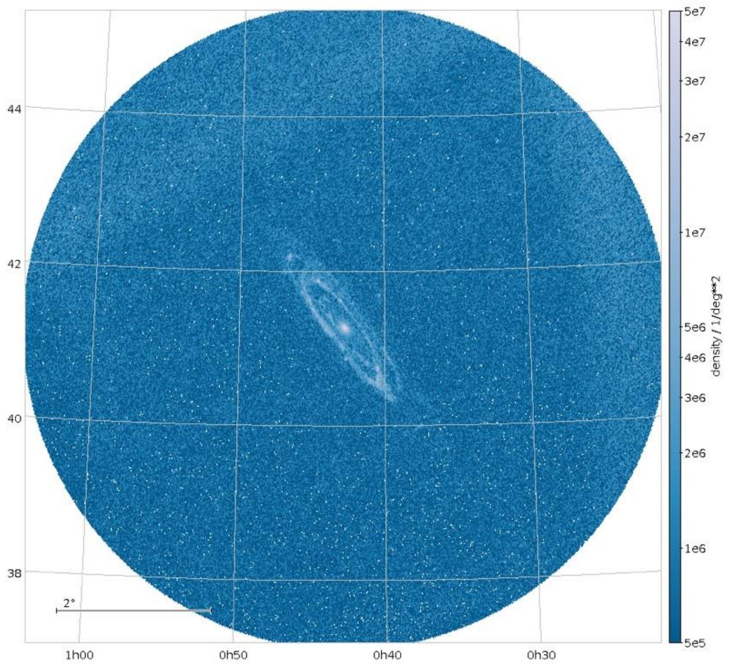
Source classifications (probabilities) plus multiple astrophysical parameters (derived from BP/RP, RVS, and astrometry) for stars, unresolved binaries, galaxies, and quasars. Some parameters may not be available for faint(er) stars.

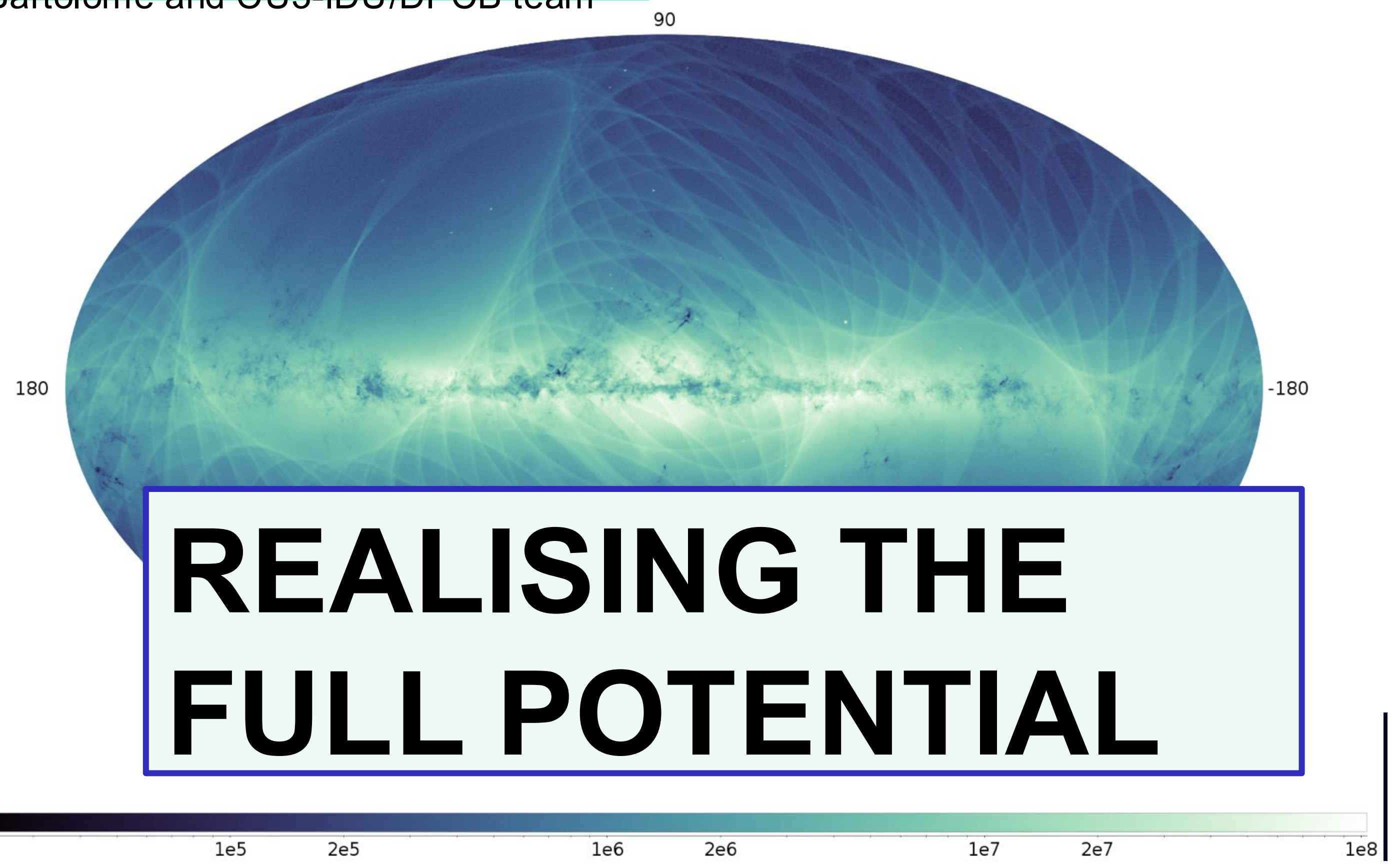
All epoch and transit (reduced) data for all sources.

Future DR5 release will include the full mission extension data

Map shows early DR4 cross match run - 142 billion detections processed – 750K CPU hours on MareNostrum Insert shows zoom in on M31

Credits: ESA/Gaia/DPAC, F. Torra, J. Portell, J. Castañeda, M. Bernet, S. Bartolomé and CU3-IDU/DPCB team

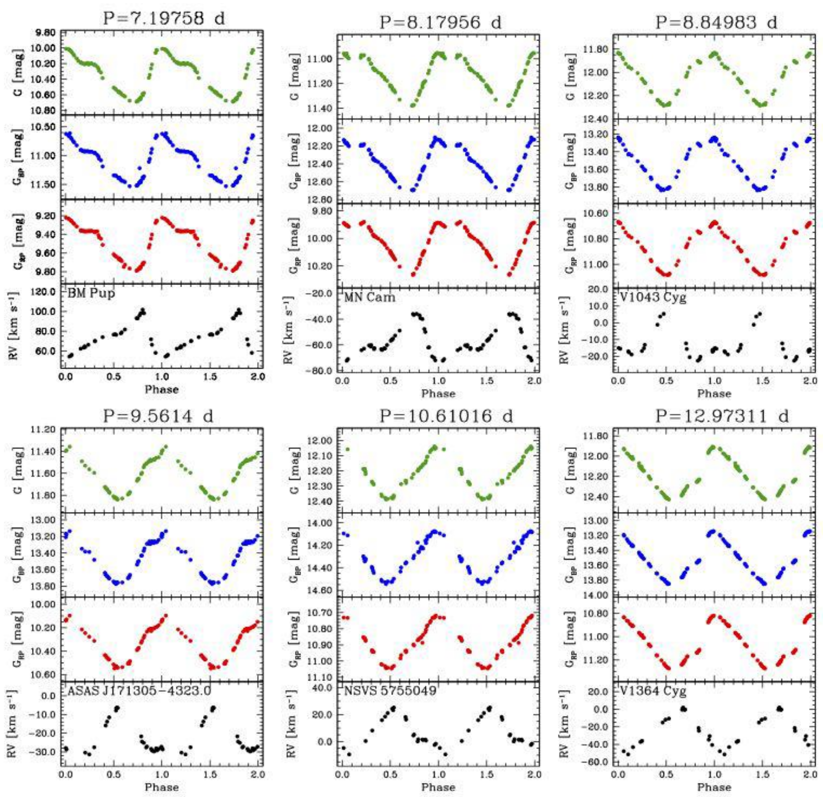


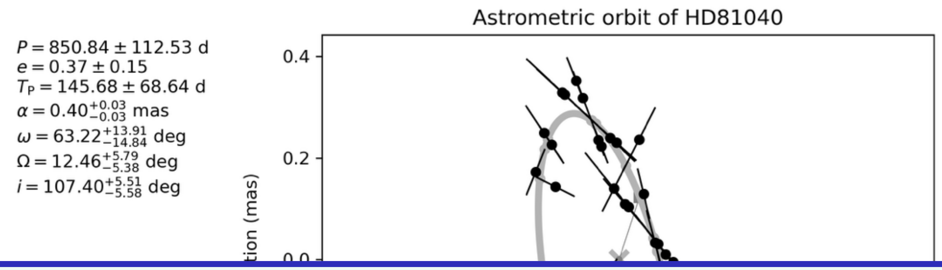


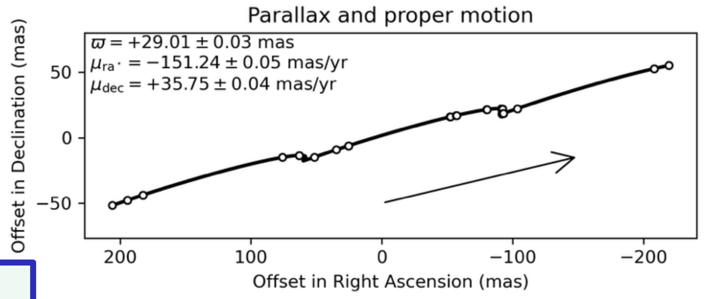
REALISING THE FULL POTENTIAL

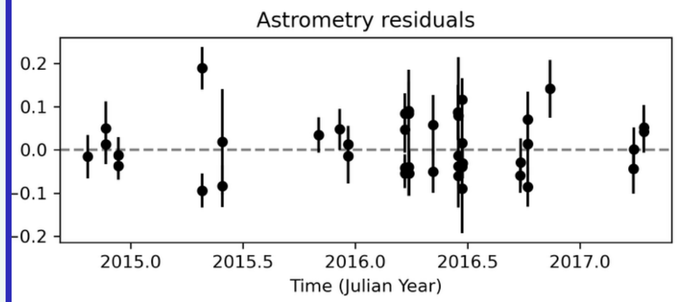
Gaia DR4: Epoch data for everything!

Epoch astrometry
Epoch light curves
Rad vel time series
XP spectra time series
RVS spectra time series

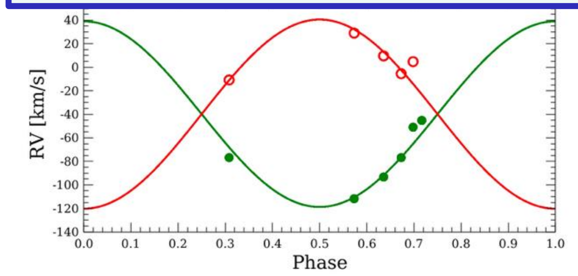


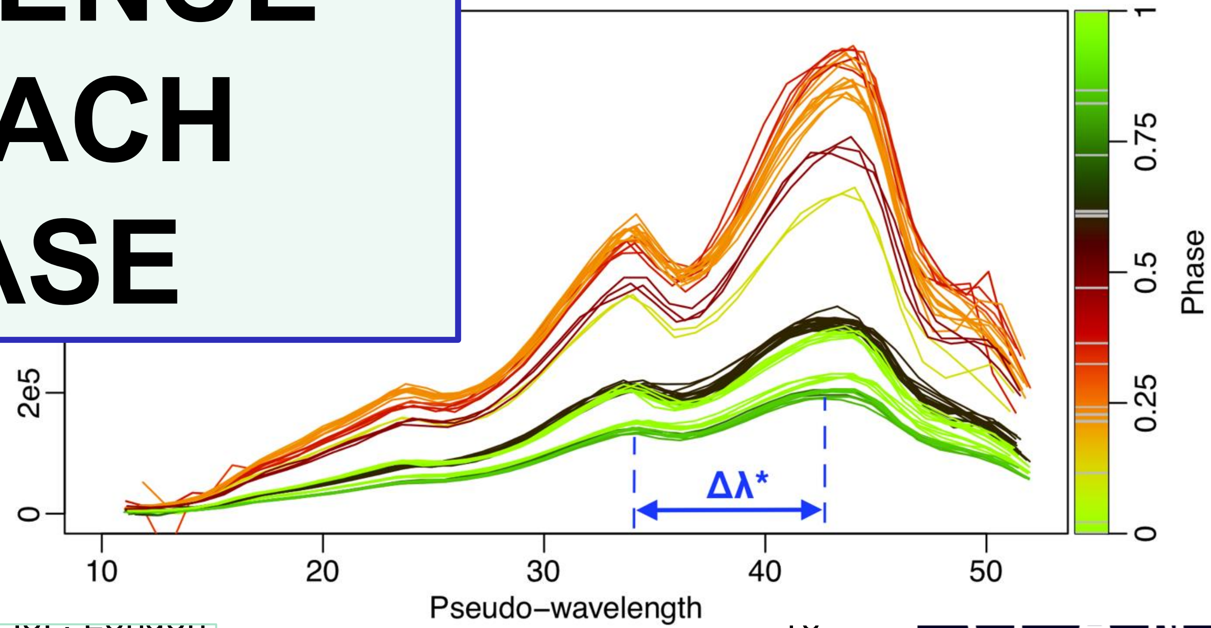






REALISING THE FULL POTENTIAL, NEW SCIENCE WITH EACH RELEASE





Nic Walton @ IoA, London

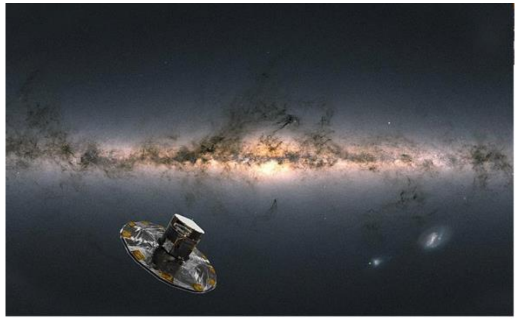
All images: credit: ESA/ Gaia/ DPAC

SPACE ACADEMIC NETWORK

Independent report highlights impact of UK contribution to the Gaia mission

https://www.gov.uk/government/news/report-highlights-impact-of-uk-contribution-to-the-gaia-mission

published: 10 Oct 2023



Gaia Interim Impact Evaluation

Interim Impact Report for the UK SPACE AGENCY

know.space

March 2023

“There can be little doubt that **Gaia is a scientific success story**. Our report highlights how UK funding has enabled the processing and release of high-quality data that is of huge value to the global scientific community, helping to advance the frontiers of human knowledge. The impact is also set to continue for years and decades to come.”

“Our analysis also shows how funding has helped to **expand the reach and reputation of the UK space sector**, develop sought-after data skills, inspire the next generation, and lead to **widespread ‘spillover’ benefits** as new advanced statistical methods, algorithms and capabilities are applied elsewhere in the space sector and beyond.”

Excellent Value for Money

IMPACT

Figure 5 *Publications by year: total publications and UK involvement*

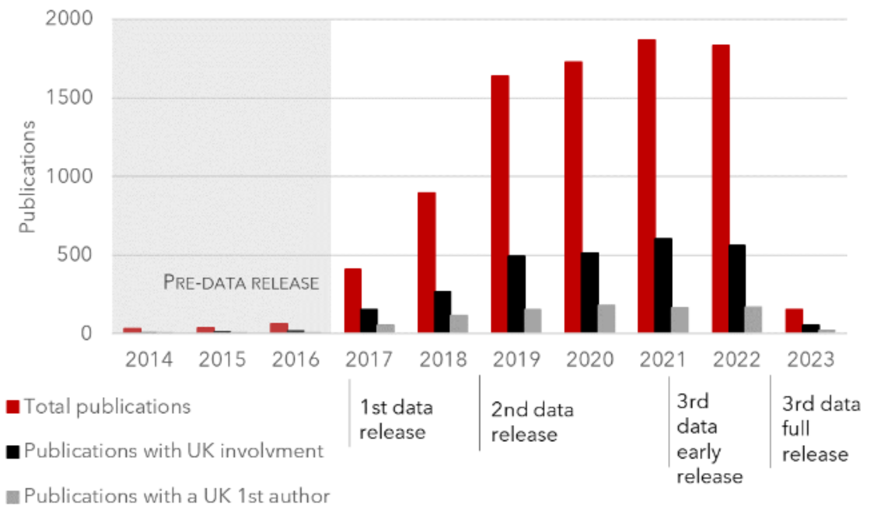
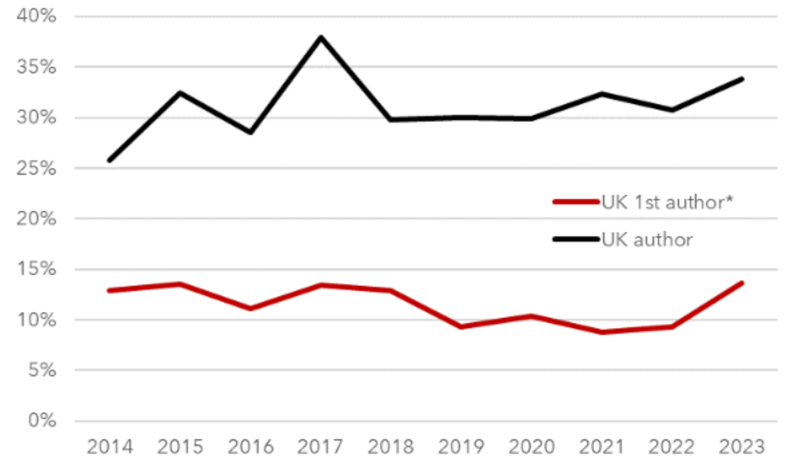


Figure 6 *Proportion of publications with a UK author or first (corresponding) author*



Papers, Science, Impact: value for money for the UK

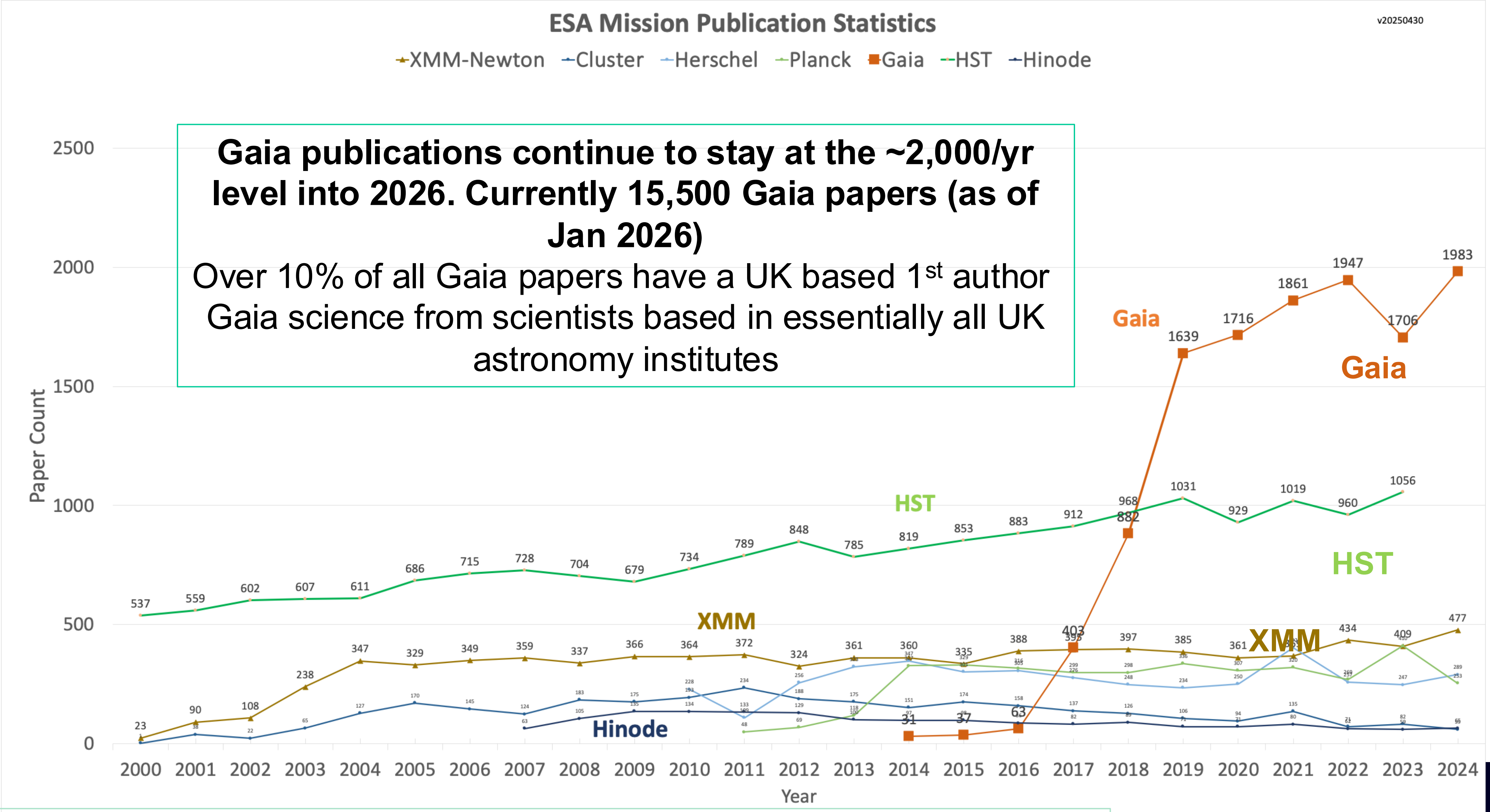
Gaia delivers for the community

SCIENCE IMPACT

Gaia publications continue to stay at the ~2,000/yr level into 2026. Currently 15,500 Gaia papers (as of Jan 2026)

Over 10% of all Gaia papers have a UK based 1st author

Gaia science from scientists based in essentially all UK astronomy institutes



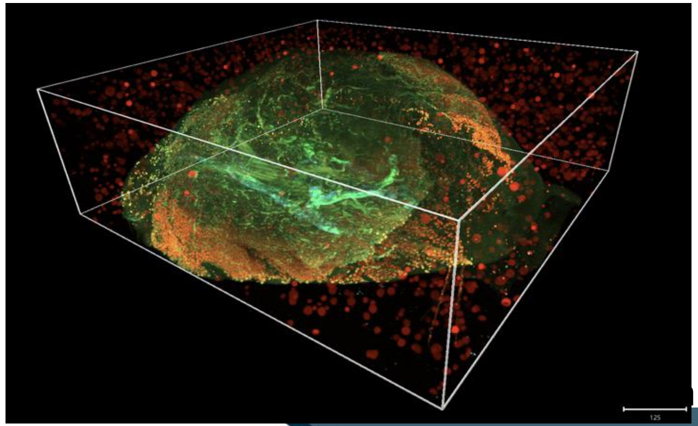
See: arXiv:2402.12818v1 and https://www.cosmos.esa.int/web/gaia/peer-reviewed-journals

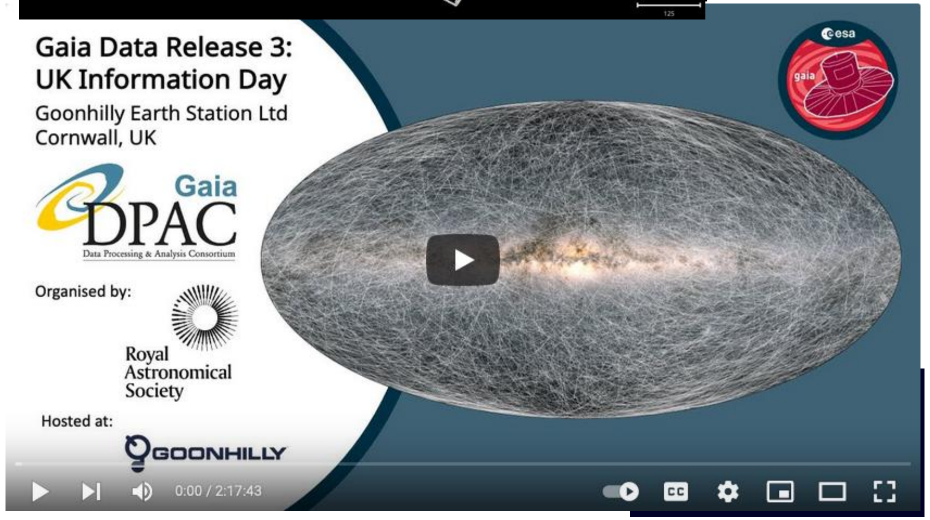
PAN
SPACE ACADEMIC NETWORK

Training, Impact, Skills
Gaia UK delivers

ECONOMIC, SOCIETAL IMPACT

- Training: Gaia underpins a significant fraction of UK astronomy research projects
  - Gaia is the most requested facility of STFC AAP supported science
  - Most UK astronomy groups (33) involved in Gaia EDR3 and DR3 related papers
- Knowledge Transfer: building on Gaia 'Big Data' techniques
  - Leveraging expertise in large scale distributed processing
  - Training via STFC PhD CDTs
  - Techniques used in medical imaging
- Outreach
  - Workshops at the RAS NAM meetings
  - Training through wider EU networks





SPANS’s Space Data White Paper

This paper generated to review space data state of the art and areas where policy and investment can underpin the UK ambitions in Space

SPAN
SPACE ACADEMIC NETWORK

UK Space Data: A World-Leading Portfolio

Space Science	JWST (MIRI lead), Gaia (DPAC), Solar Orbiter, Euclid (VIS lead), PLATO, ARIEL (UK lead), JUICE, LISA, BepiColombo, ExoMars	UK lead roles on multiple ESA space science missions
Earth Observation	Copernicus/Sentinel, Biomass, FORUM, MicroCarb, NovaSAR	
Space Domain Awareness	Met Office MOSWOC, Solar Orbiter MAG, Vigil (L5), BGS observatories, Chilbolton, NSpOC	

The UK has deep involvement in 20+ current space missions with significant data processing responsibilities

SPAN
SPACE ACADEMIC NETWORK

UK Data Infrastructure Landscape

Mission Data Processing Centres	Cambridge (Gaia), Edinburgh (Euclid), UCL/MSSL (SolOrb), Leicester (Bepi)
IRIS	Coordinated computing, storage & networking for STFC science communities
DiRAC	Distributed HPC across 4 universities. Co-designs hardware for science
AIRR	AI Research Resource: Isambard-AI (Bristol) & Dawn (Cambridge). Pre-exascale
CEDA / JASMIN	EO & atmospheric data. Data-intensive computing for environmental science
EO Data Hub	UK Pathfinder for EO data access. Led by NCEO. Low-TRL exploration
ESA ESDC / ESO SAF	European archives: full range of ESA missions + ground-based data at the ESO Science Archive

But: No UKSA data policy. Unclear agreements on cross council access to facilities (e.g. UKSA use of IRIS)

SPAN
SPACE ACADEMIC NETWORK

Example White Paper Input

Astronet Roadmap

A Computing; big data, HPC and data infrastructure

Top Level Key recommendations

Key 1: People

[People] we recommend developing and investing in a professional software engineering / computational skills base in Astronomy. This has many implications and requirements, including career development with clear progression pathways in academia and improving the diversity of the workforce. Such careers have to be promoted and considered as an integral part of our science/research portfolio. Traditional metrics for academic performance are often inappropriate for measuring the impact and usefulness of computationally focussed outputs. New assessment criteria, for example based on industrial models, should be adopted by the astronomy community to give proper credit to essential contributions that technicians and software engineers provide.

Key 2: Life cycle planning

[Data, Open, Archive, User-to-Data] we recommend that missions and facilities plan an integrated approach for data products and software tools: their design, delivery, maintenance and development should be sufficiently planned for and resourced already at the onset and for the lifetime of the mission/facility. We recommend that initiatives be supported for the long term preservation and scientific use of data.

Key 3: ‘Tiered’ data architecture

[Data, Archive, User-to-Data, Open, Green] we recommend to adopt and further develop a “Tiered” approach for Data Infrastructure, for all types of data pertaining to astrophysics, including models, simulations and mocks, and where beneficial to connect with similar frameworks developed for other disciplines of science. A `Tiered' Distributed Analysis model maximises the transformation of science data to scientific results, organising and sharing expertise to benefit from economies of scale, with attention given to preventing the exclusion of individual communities, as well as monitoring and minimising the environmental cost.

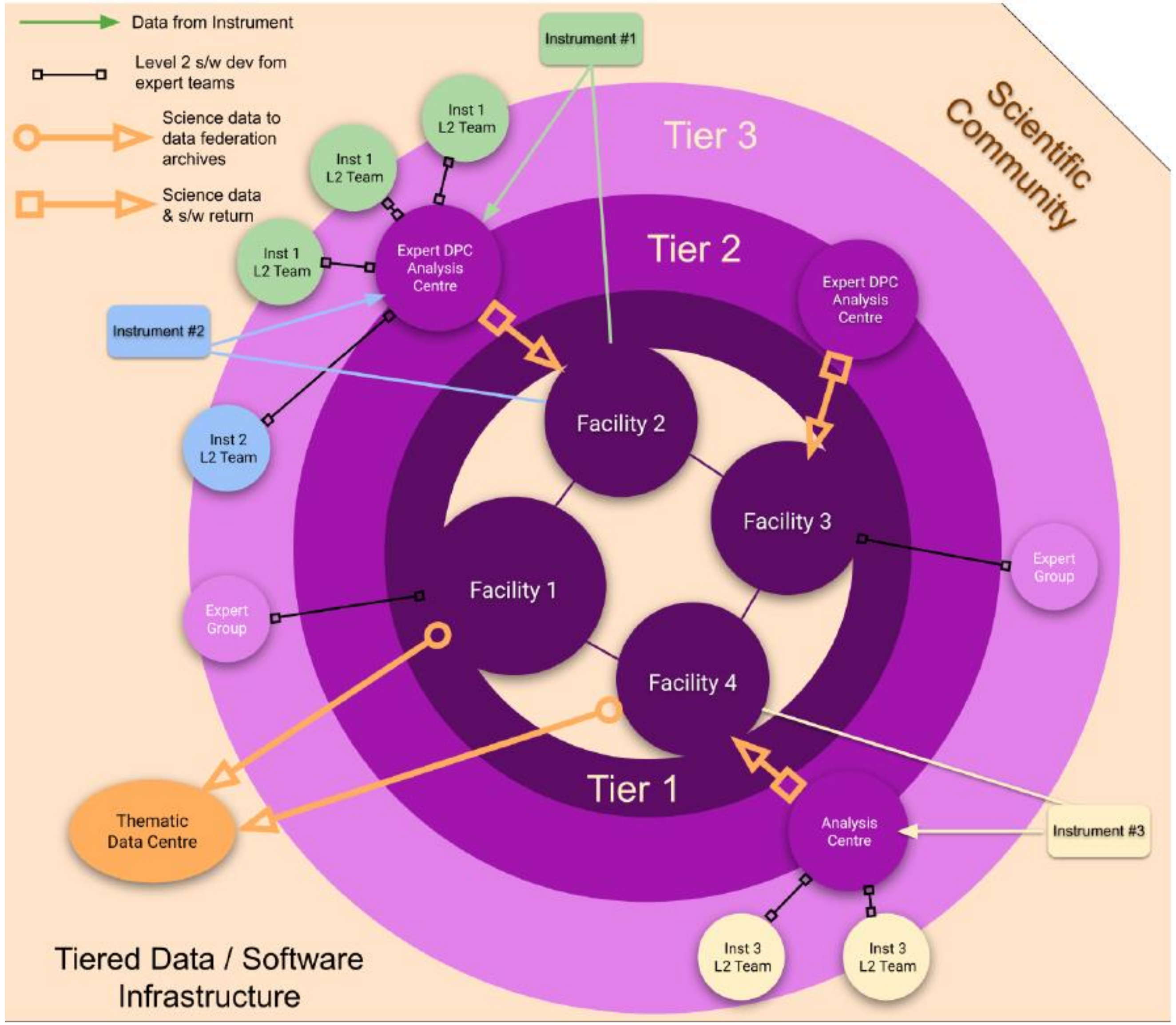
Key 4 : Open data

[Open, User-to-Data, People, Green] we recommend embracing a fully collaborative, open and synergistic view when it comes to the astronomy-computing ecosystem, encompassing data, software, processing, analysing and modelling. The community should implement and support official mechanisms to monitor, acknowledge and reward behaviours that model this ambition.

Open science, data and software sharing, archives, cloud computing, platforms and service infrastructure represent various facets of an integrated view of computing in astronomy: this should be acknowledged and acted upon.

Keywords in brackets refer to specific sections of this document:

- [Open] = The paradigm change: Open Science, Data and Software sharing
- [Data] = Mission / Facility data processing
- [Archive] = Archives for data and software



SPAN SPACE ACADEMIC NETWORK

Critical Challenges: Pipelines & Software
Generating the underpinning data products

- Funding (can) prioritise hardware over software investment
- Data planning is often not planned and resourced at all stages of a space mission (from concept study through to post-operations final data releases)
- Short term nature of funding: 1 to 3yr grants vs 10–20yr mission lifetimes.
- Pipeline knowledge held by small teams, with potential loss of capacity across missions

Critical Challenges: Coordination

- The UK Space Agency currently lacks a clear data policy comparable to the more joined-up approaches adopted by organisations such as NERC (e.g. EO Data Hub, EDS).
- The lack of unified national space data infrastructure means that UK-held space data is fragmented across multiple repositories with no single discovery portal.
- Space data has no defined place within the emerging national data library framework, representing both a gap and an opportunity for strategic engagement.

SPAN
SPACE ACADEMIC NETWORK

Critical Challenges: Skills and Capacity

- Skills gaps in analysis pipeline development, advanced data analytics and AI/ML applications remain a critical constraint.
- Shortage of personnel with combined domain science and software engineering expertise affects the UK's ability to develop and maintain data pipelines, exploit new analysis techniques, and maximise scientific return from space data investments.
- Limited structured career paths for Research Software Engineers, especially in non-industrial settings (e.g. Academia)
- Recruitment issues, many space organisations struggle to recruit appropriately skilled data staff

SPAN
SPACE ACADEMIC NETWORK

Critical Challenges: Data Resilience

- Space recognised as Critical National Infrastructure
- Critical dependencies on external data (e.g. NASA/NOAA) and commercial cloud platforms (e.g. AWS, Google Cloud)
- Potential reductions in US science budgets threaten access to key data archives (e.g. NASA Planetary Data System and Astrophysics Data System)
- Key infrastructure (e.g. Sentinel Hub) owned by non-UK entities
- National Space Operations Centre will depend on robust data architectures connecting ground sensors, space weather monitoring, and orbital tracking

Critical Challenges: AI, Cloud & Emerging Risks

- Cloud platforms may potentially scrape data for LLM training: data sovereignty, licensing and IP concerns
- EU AI Act classifies AI systems by risk level: EO systems used for critical infrastructure monitoring may fall within regulated categories
- AI trustworthiness, validation, and reproducibility challenges for scientific applications
- International standards for AI in space applications are in their early stages: UK should contribute to shaping them
- Satellite mega-constellations impacting astronomical observations

Policy Context: Timelines

- Nov 2025: ESA budget: latest CMIN25 saw a £1.7B commitment from the UK (now 5th largest funder of ESA after Germany., France, Italy and Spain)
- Mar 2026: Significant UK Gov announcement on UK Space funding investment (£500M) in the period 2026-2030
- Apr 2026: Restructuring of 'Space' coordination with UK Space Agency joining DSIT ('reducing bureaucracy', 'cutting duplication')
- Late 2026: 'Plan for Space' to be publish by UK Gov

The SPAN Data White Paper can inform the Plan for Space and the UKRI's ongoing Infrastructure Roadmap refresh

SPAN
SPACE ACADEMIC NETWORK

The International Policy Context

- NASA SPD-41 Science Implementation Policy: mandates data management plans for all science missions from proposal stage: the UK has no equivalent
- ESA EarthCODE: requires FAIR compliance for space data
- EOSC EU Node: launched Oct 2024, first node of European research data federation
- EU AI Act: classifies AI by risk level with implications for EO in regulated contexts
- ESA Voyage 2050: 3 Large mission themes to 2050+, each at unprecedented data scales
- Astronet Roadmap: data infrastructure equally important as instrumentation

SPAN
SPACE ACADEMIC NETWORK

Ten Priority Actions

#	Action	Priority	Cost Band
1	1.2	Develop a UK Space Data Policy	LOW
2	7.1	Space Data Coordination Within DSIT	LOW
3	4.1	Assess International Data Dependencies	LOW
4	1.1	Pipeline Planning in Mission Design (incl. calibration/QA)	LOW
5	2.1/2.2	Prepare for Next-Generation Missions & Surveys	LOW+MED
6	1.4/3.3	Strengthen Pipeline Expertise & Career Pathways	LOW+HIGH
7	1.3	Establish UK Space Data Discovery Service	MED
8	4.3	Engage with European Open Science Cloud	MED
9	5.1/5.3	Address Environmental Sustainability	LOW+MED
10	7.2	Sustainable Funding for Coordination & Roadmap	LOW

Aligning with Current Space Priorities

Space Data Theme	SatComms*	Access*	ISAM*	SDA*	EO	PNT	Space Science
8.1 Pipelines	●	●	●	●	○	○	○
8.2 Next-Gen Missions				●	○		○
8.3 Skills	●	●	●	●	○	○	○
8.4 Intl/Resil				●	○	○	○
8.5 Sustainability	●	●	●	●	○	○	○
8.6 Collab	●		●		○		○
8.7 Governance	●	●	●	●	○	○	○
8.8 Delivery	●	●	●	●	○	○	○

● = serves priority subsector ○ = serves non-priority subsector * = Government priority [30]

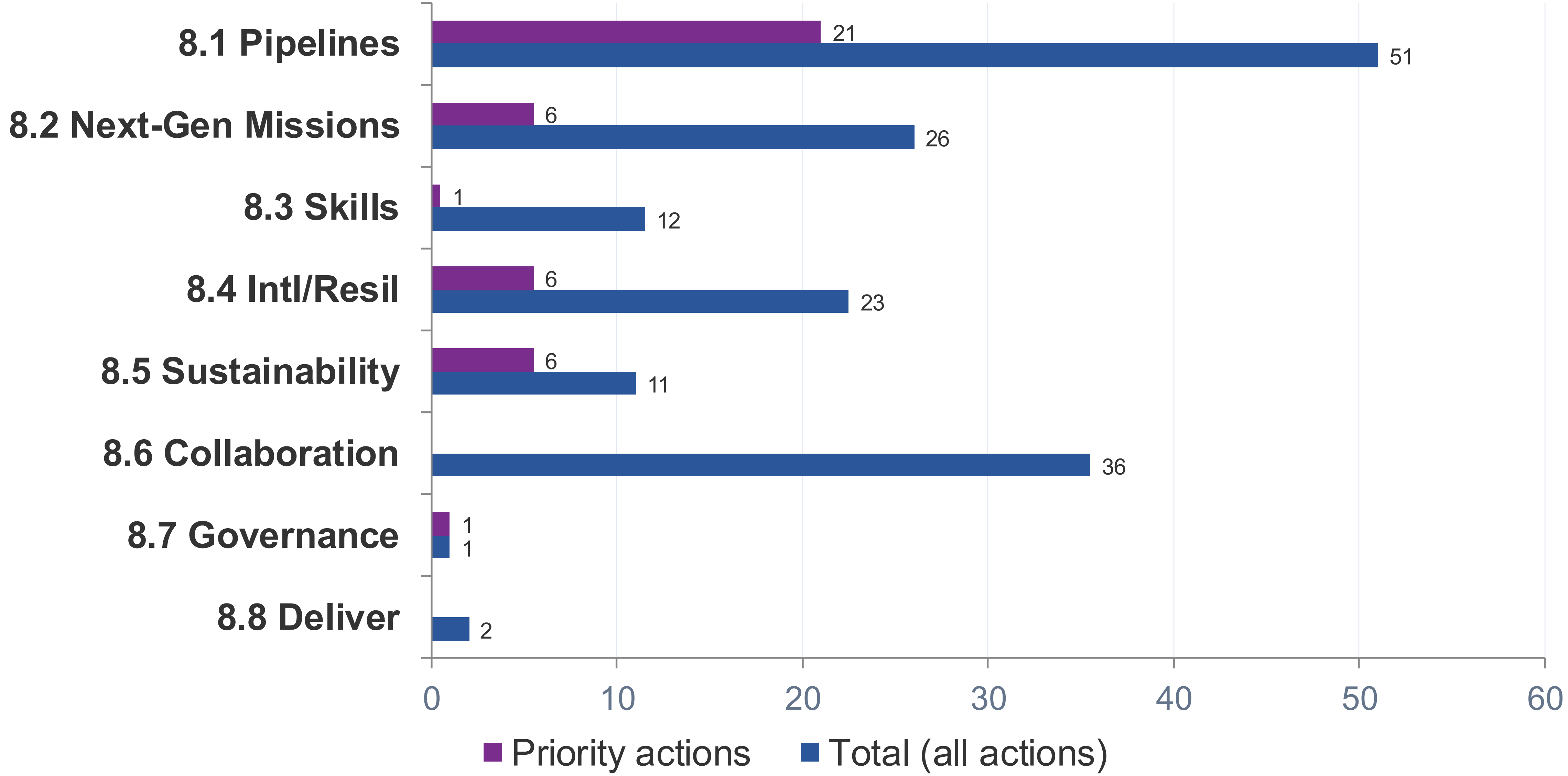
SPAN
SPACE ACADEMIC NETWORK

Action Types and Resource Ranges

Band	Cost (over 3 to 5 year timeframe)	Nature of Investment
LOW	< £1M	Policy, reviews, assessments, coordination. Achievable through reprioritisation of existing budgets or modest additional funding.
MEDIUM	£1M – £10M	Service development, pilot programmes, fellowship schemes, standards engagement. Requires dedicated programme funding but within scale of existing UKSA/STFC call mechanisms (e.g. NSIP, STFC grants).
HIGH	> £10M	Infrastructure build, sustained multi-year programmes, large-scale computing/storage. These will require a spending review case or co-investment with national infrastructures e.g. IRIS, JASMIN, AIRR, SCIF.

SPAN
SPACE ACADEMIC NETWORK

ROM Implementation Costs



Time for Action: Implementation in Phases

Initial Phase: from year 1:

- All 10 priority topics. Data Policy, DSIT coordination, dependencies, pipeline planning, next-gen missions, RSE workforce, Discovery Service, EOSC, environment, roadmap.

Phase 2: from years 2-3:

- Everything above plus: fellowships, apprenticeships, science-ready data services, ML/AI, open ecosystem, multi-messenger, real-time pathways, equitable access.

Phase 3: from year 3-4 onwards:

- Everything above plus: tiered infrastructure, resilience capability, regional clusters, co-location labs. 'High' items can leverage investment in related programmes, co-investment across DSIT agencies, etc.

SPAN
SPACE ACADEMIC NETWORK

The Role of SPAN

SPAN's role: complement and inform DSIT and UKRI and other relevant bodies, agencies.

- Independent analysis and community consultation
- Cross-mission perspective no single institution can offer
- Technical review to support government decision-making
- The collective academic voice in space data policy

Next Steps: Short term (1-2 years)

- Submit whitepaper to DSIT as input to the Plan for Space (2026)
- Engage new DSIT space unit on data infrastructure oversight function
- Support development of UK Space Data Policy through SDAWG expertise
- Position UK space data within UKRI Infrastructure Roadmap refresh
- Contribute academic perspective to spending review discussions for data infrastructure investment
- Develop spending review case for HIGH-cost infrastructure items

SPAN
SPACE ACADEMIC NETWORK

Next Steps: Longer term (3-5 years)

- Establish UK Space Data Policy by 2027
- Implement tiered processing infrastructure leveraging IRIS, JASMIN, DiRAC, AIRR
- Support ESA Voyage 2050 data planning (e.g. for L4 Enceladus, L5 GaiaNIR)
- Build Discovery Service prototype
    - connecting space science, EO and SDA data
- Connect to EOSC federation; engage with EU frameworks (AI Act, Data Act, Space Act)
- Develop environmental impact reduction plans for UK space data infrastructure

Space for Discussion

- Discussion now from first look at the challenges and actions
  - Feedback on priorities, coverage (what's missing), implementation possibilities, audience and consumers for the White Paper
- This White Paper should be update in light of community feedback
  - SPAN will update the document to reflect feedback
- The document will be reissued considering progress made against the action plan
- The look ahead for the White Paper currently runs for the upcoming five years through to 2031
  - Within this period there will be development of DSIT plans for Space for the following spending review period (from 2030), where inputs to strengthen Space Data are timely

SPAN
SPACE ACADEMIC NETWORK

Summary

- Embed data infrastructure in the Plan for Space as a cross-cutting enabling framework to support wider UK Space priorities
- Adopt the 10 priority actions for near term implementation
  - 7 of 10 are ‘Low’ cost, policy changes, not capital spend
- Ensure the new DSIT UKSA unit has data infrastructure oversight
  - coordinate data investment alongside hardware and mission commitments
- Invest in people: the pipeline workforce is the binding constraint
  - Underpin longer term RSE posts, career pathways, dual-sector fellowships
- Recognise data as the mechanism through which hardware investment generates scientific, economic and societal return

SPAN
SPACE ACADEMIC NETWORK

SPAN SPAN Data White Paper:

Presentation slides

The SPAN white paper ‘White Paper on Space Data Excellence in Growth: Space Data Infrastructure for a Competitive UK’ is available for download from https://span.ac.uk/publications/

at:

https://span.ac.uk/wp-content/uploads/2026/03/span-data-whitepaper-v1.0.pdf