

IRIS Capacity at a Turning Point

Andrew Sansum

20/10/26



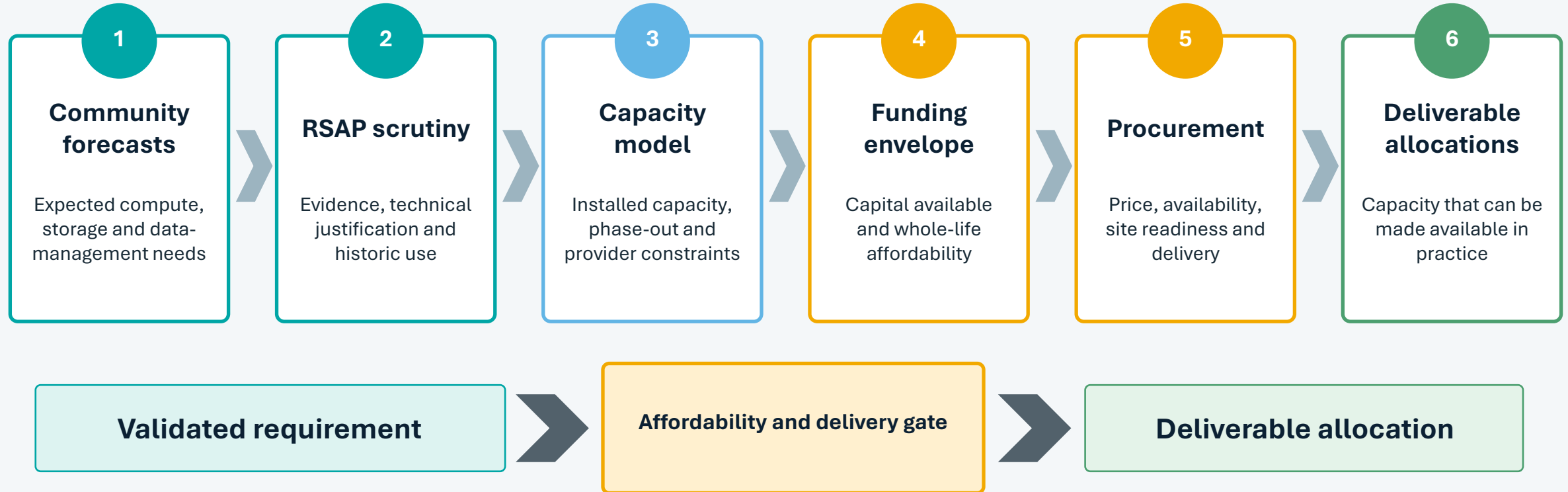
Science and
Technology
Facilities Council

Scientific Computing



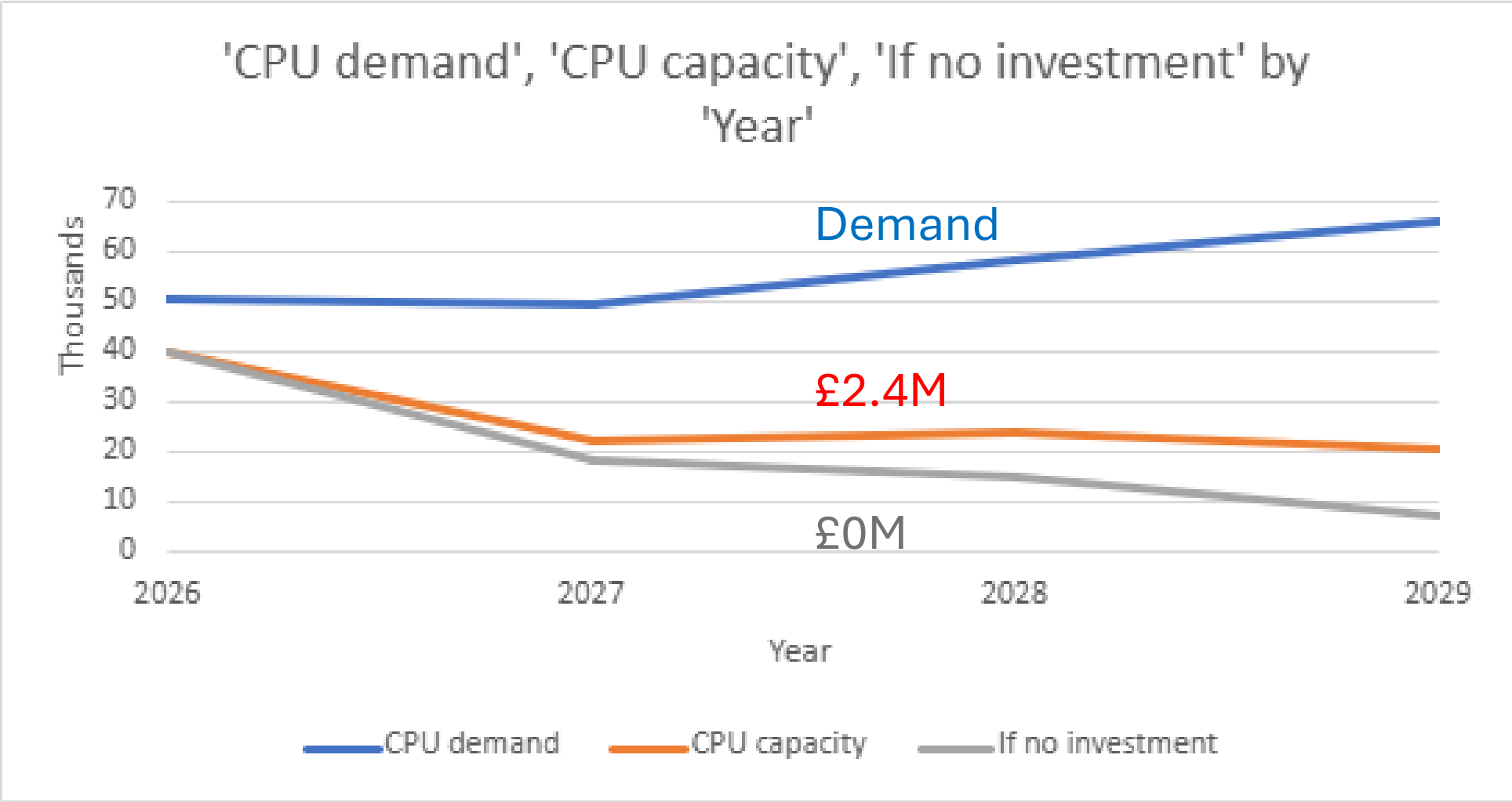
The planning chain

The process separates evidence of need from decisions about affordability and delivery.

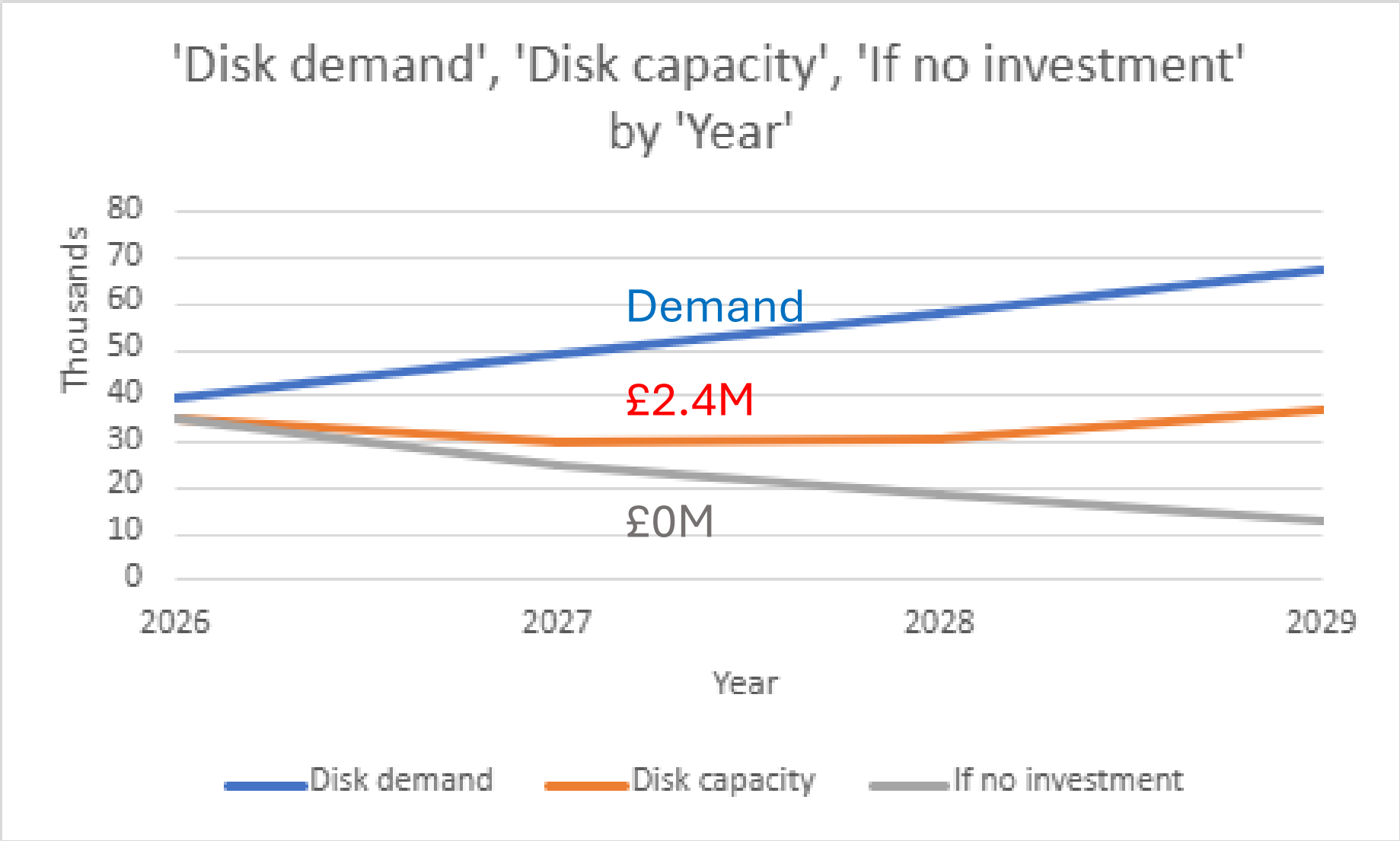


The shortfall is created after technical scrutiny: justified demand can exceed affordable, procurable capacity.

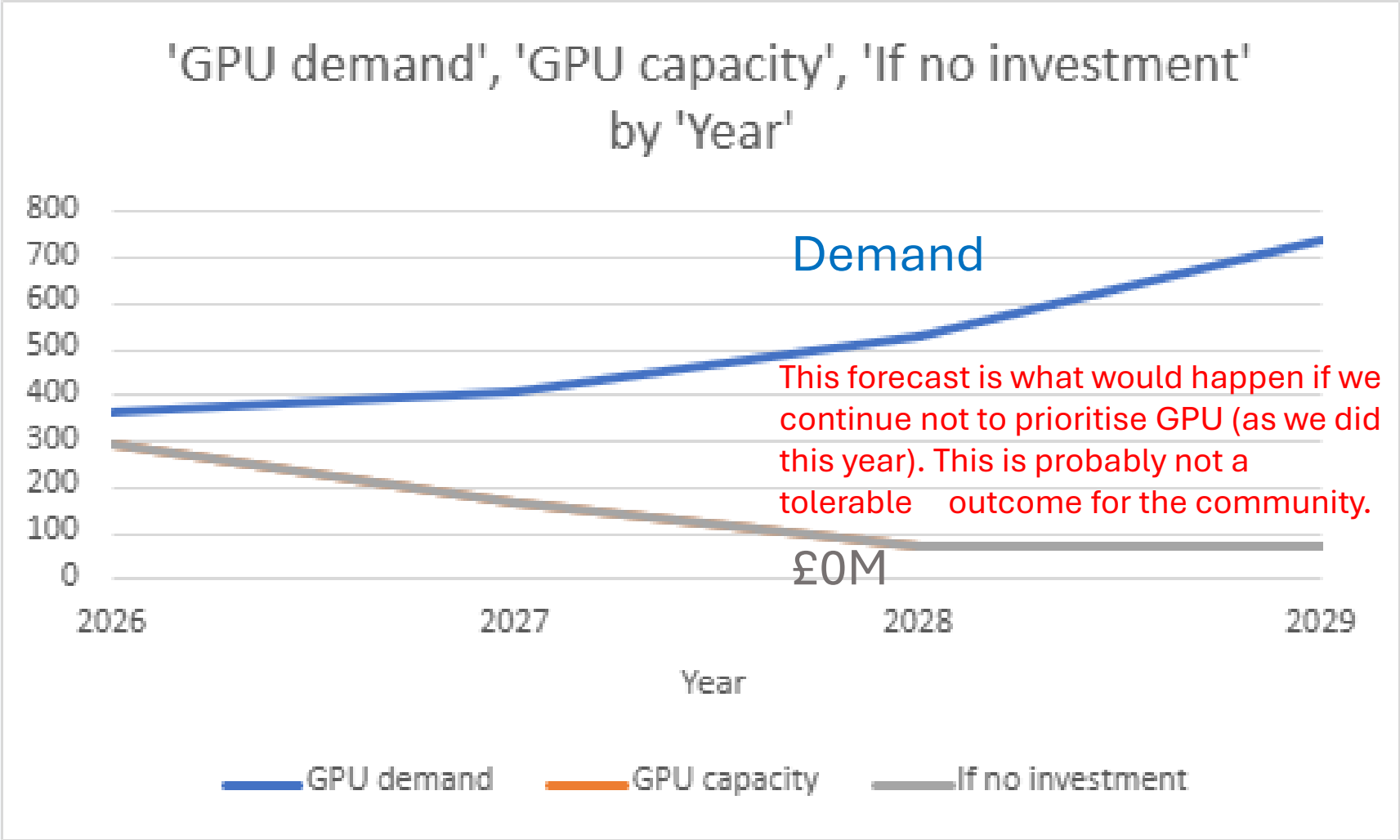
CPU Capacity Model: £2.4M (2027-2029)



Disk Capacity Model: £2.4M (2027-2029)



GPU Capacity Model: £2.4M (2027-2029)



Capacity falls before growth is considered

The replacement requirement is concentrated because earlier hardware purchases were uneven.

HARDWARE LIFECYCLE



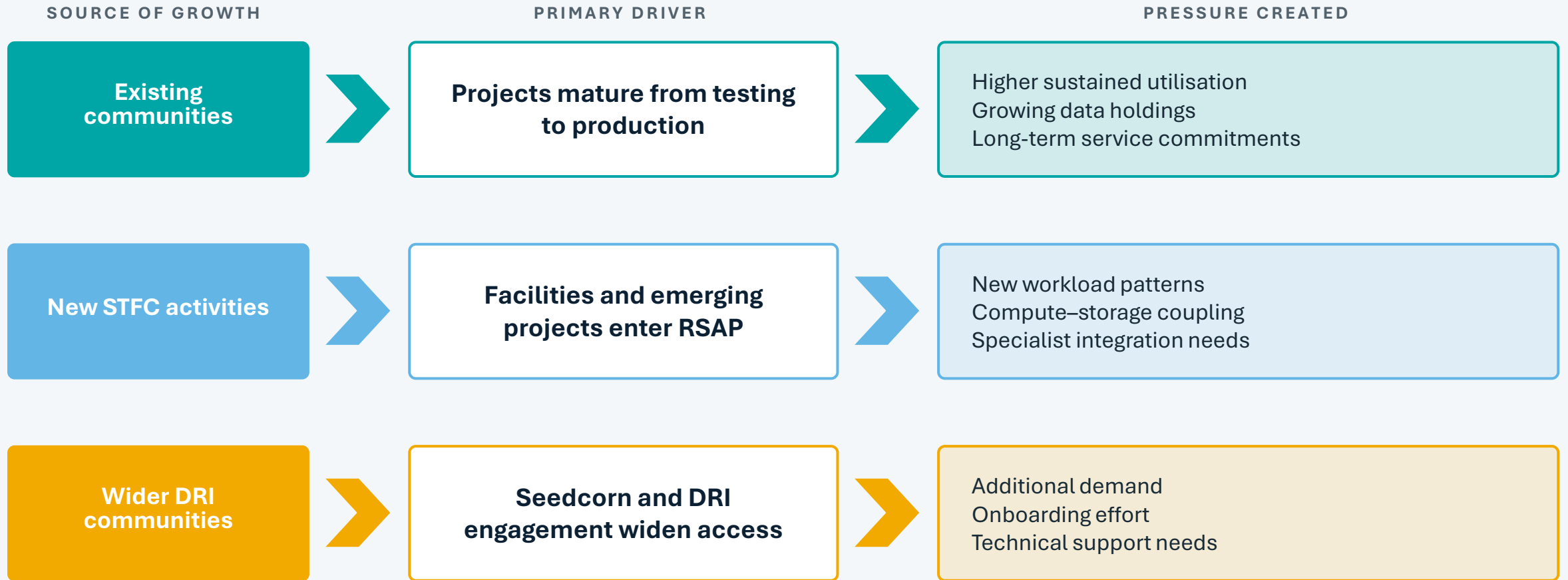
WHAT THIS MEANS FOR AVAILABLE CAPACITY



Extending asset life can defer capital expenditure, but it transfers pressure into maintenance, reliability, energy and provider support.

Demand is growing from three directions

Each source adds a different combination of capacity, capability and support requirements.



Total demand growth is broader than extra CPU: storage, memory, accelerators and support all matter.

The same budget now buys less

Several cost pressures act together, reducing the capacity that can be purchased from a fixed capital envelope.

COST PRESSURES

Component prices

Pressure on memory, storage, accelerators and complete systems

Market volatility

Short quote validity, availability risk and uncertain delivery costs

Site and integration

Power, cooling, installation, commissioning and readiness

Whole-life support

Warranty, maintenance and provider operating effort

PURCHASING MODEL

Available capital

÷

Cost per usable unit

x4

=

DELIVERABLE OUTPUT

Capacity purchased

Earlier cost basis

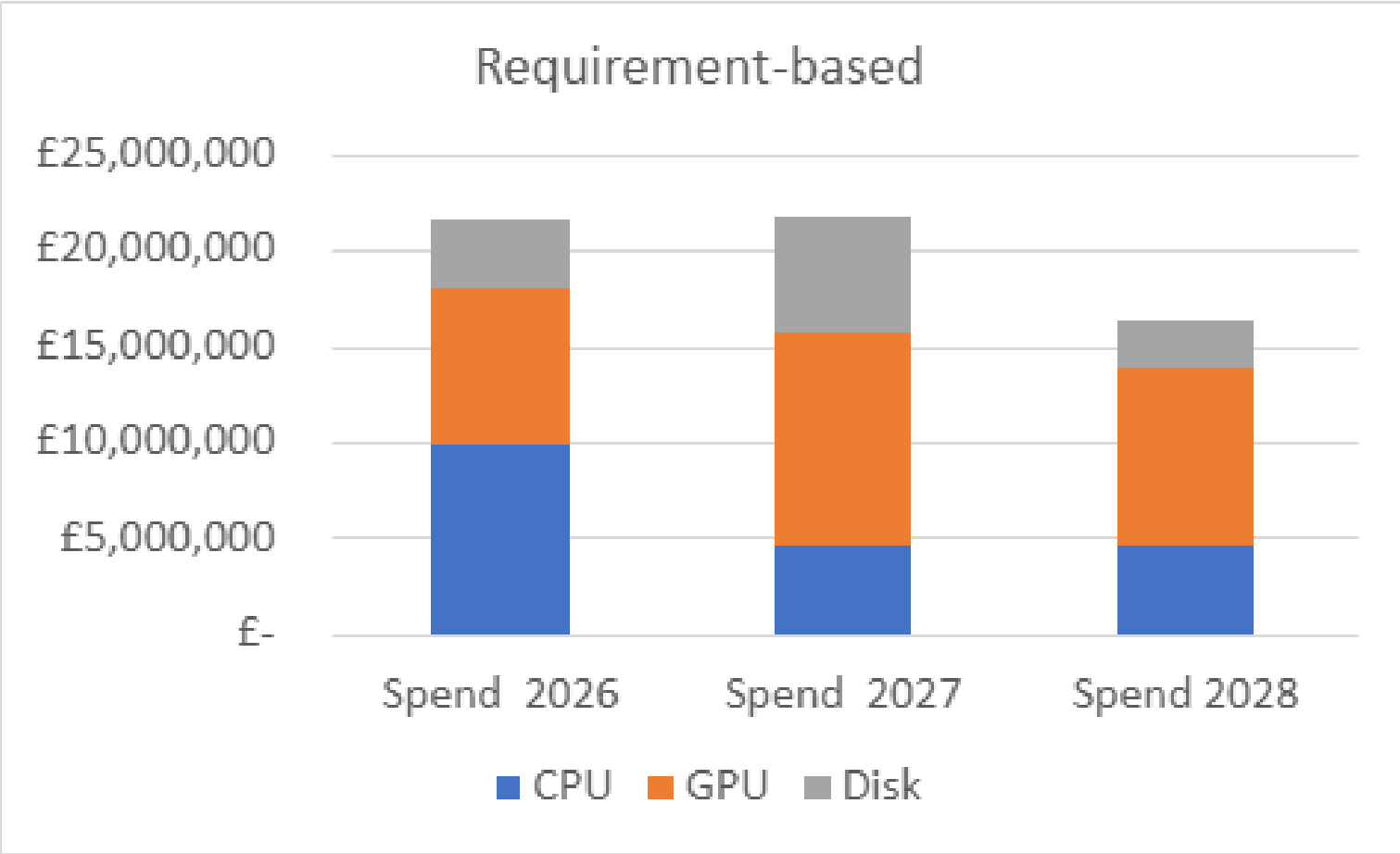
Higher hardware unit costs

Future provider M&O Costs

**Less usable capacity in future
from the same cash envelope**

Cost uncertainty is now a material planning variable: allocations must retain flexibility until procurement is complete.

Budget Requirements – August Estimate



A Strong FY26 Award, but a Much Larger Need

The question is how to protect viable services within a sharply constrained and fast-moving market.

£25.9M

tiered request
(January 2026 prices)



£4.3M

approved Exec Board (EB)
capital allocation



Service protection

not full restoration of demand

Retiring hardware

Capacity falls before new demand is considered

Growing requirements

Projects are moving from test to production

Higher unit costs

The same budget buys materially less capacity

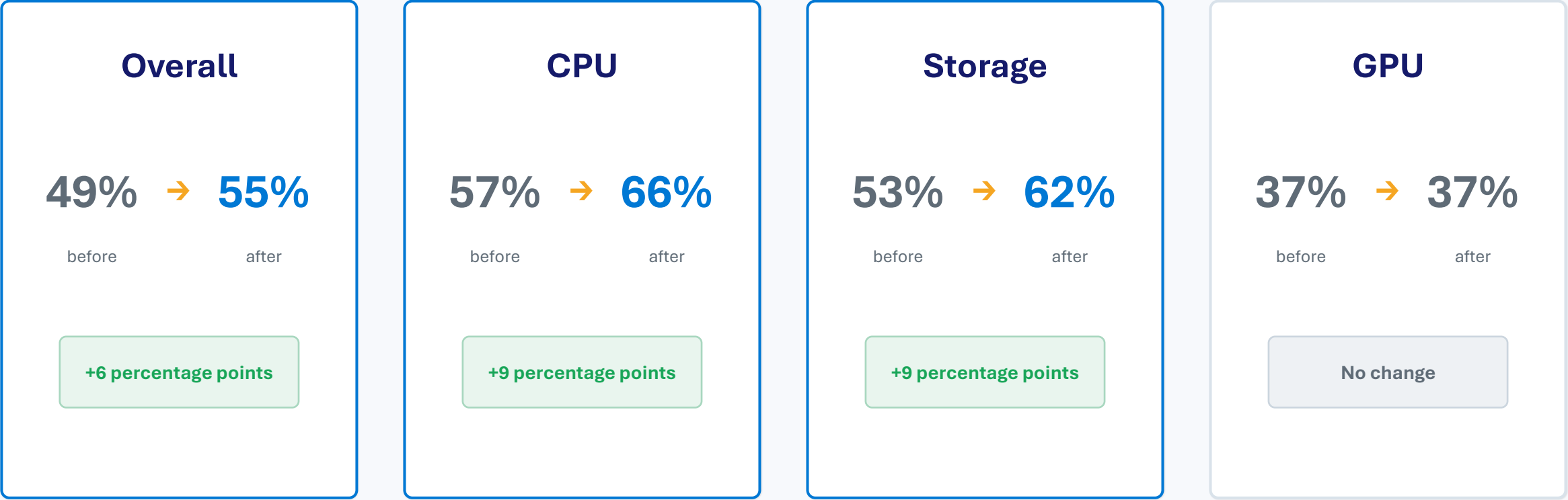
Delivery deadline

Late procurement creates underspend risk

Planning stance: prioritise operational continuity and evidenced shortfalls, then improve coverage where headroom remains.

Expected effect: meaningful risk reduction, not full coverage

Based on the June 2026 price point we could only achive the following. In practice we will achieve smaller uplifts.



Even at June prices, 45% of validated overall demand remains unmet after procurement.

Capital alone is no longer enough

Infrastructure providers increasingly need recurrent funding to operate and sustain IRIS-funded equipment.

WHY THE MODEL IS UNDER PRESSURE

Current IRIS model

IRIS funds

CAPITAL HARDWARE

No routine IRIS budget for

running costs
staff effort
maintenance

Site running costs are rising

Electricity and other operating costs increase the cost of hosting IRIS equipment.

IRIS is now a larger share of estates

Providers increasingly need IRIS to meet a fair share of costs previously absorbed by larger estates.

Engagement itself requires staff

Planning, integration, operations and engagement with IRIS consume provider staff effort.

Equipment must run for longer

Original four-year assumptions are giving way to seven years or more as replacement capital becomes harder to afford.

Longer life creates new costs

Extended operation can reduce residual provider benefit and require longer maintenance and support contracts.

Other funded effort is shrinking

STFC prioritisation may reduce effort funded by other communities, leaving IRIS to cover more of the operational burden.

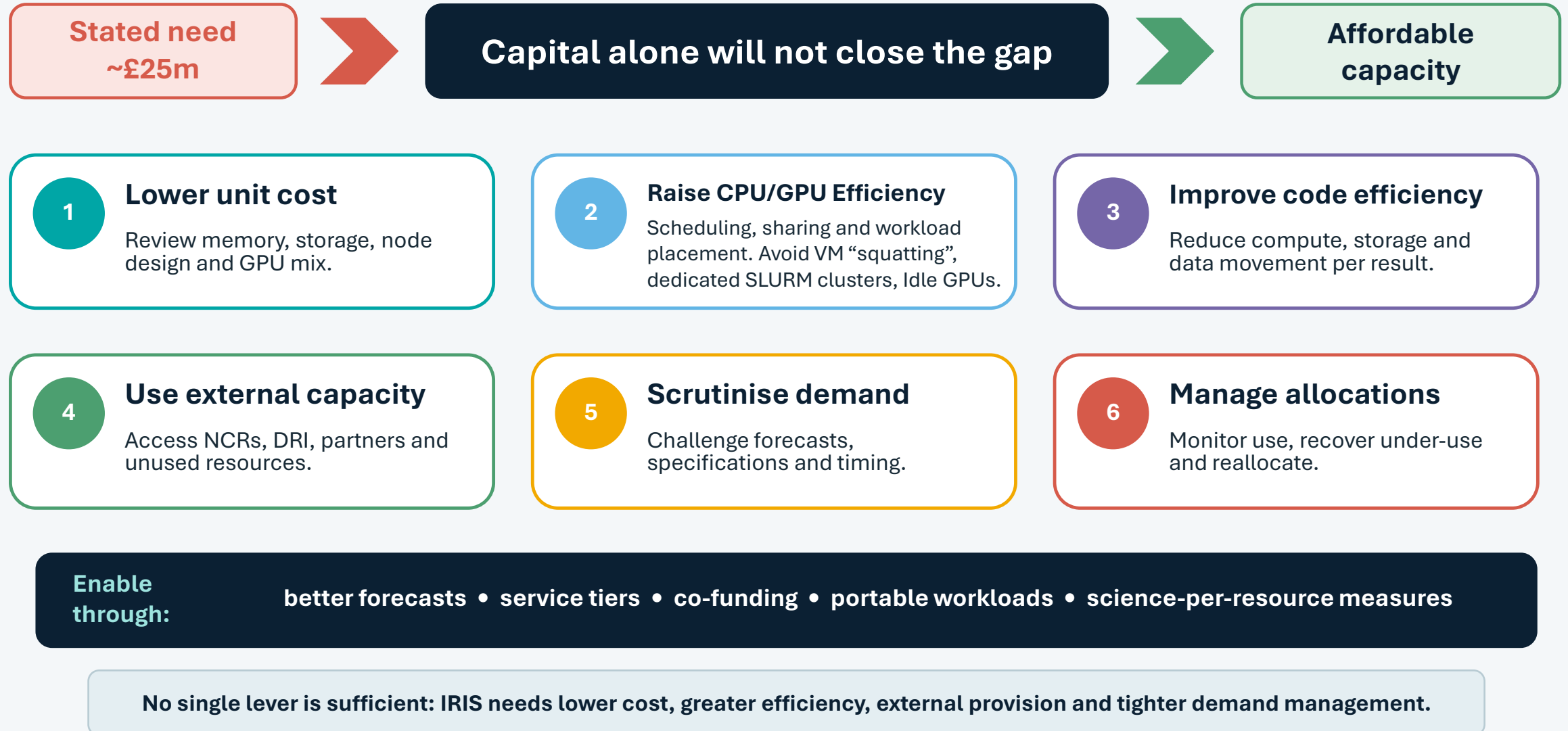
Funding question

Can IRIS convert some capital into recurrent funding, or establish a model that explicitly pays provider staff effort, maintenance and running costs?

The issue is structural: sustaining capacity increasingly requires both capital and recurrent resource.

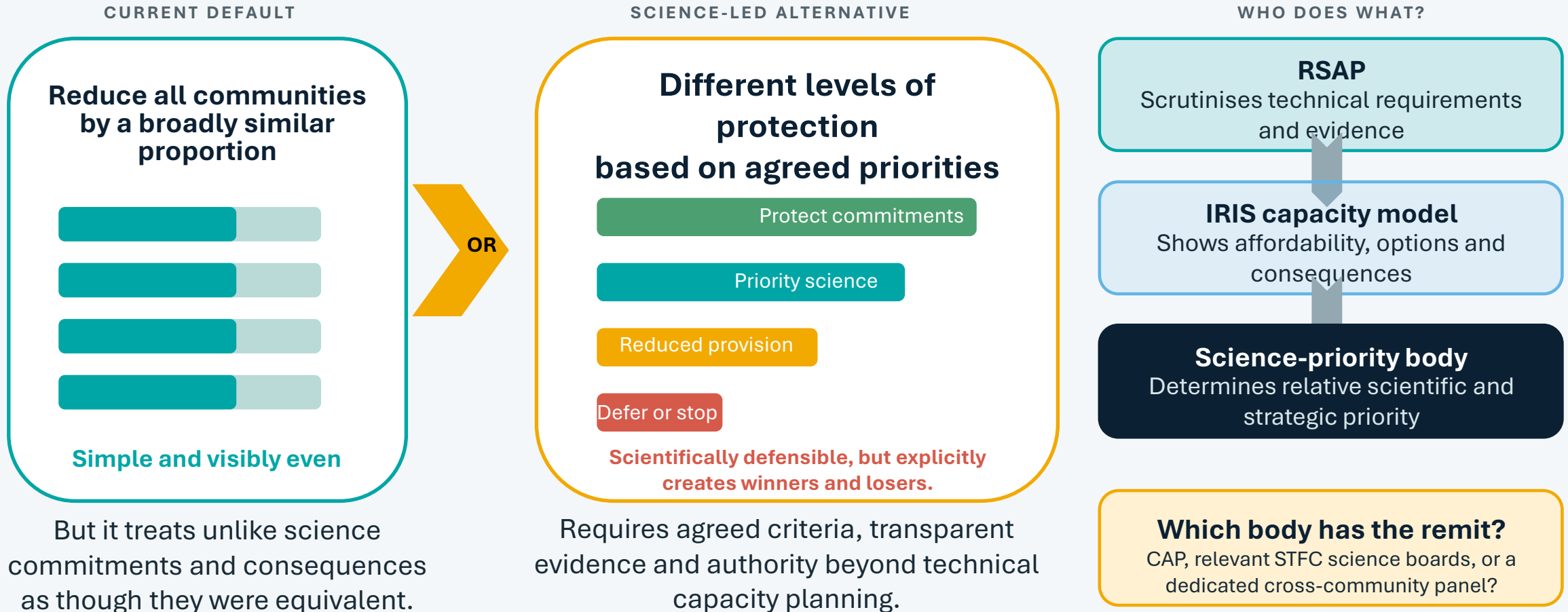
Closing the Gap

A £25m capital budget is unrealistic. IRIS must use several complementary levers.



The Shortfall Changes the Nature of the Decision

IRIS is moving from capacity planning into explicit prioritisation.



Decision required: retain broadly equal reductions, or establish a science-led process with clear criteria, accountability and an authorised decision body.

The Central Message

IRIS faces a structural gap between validated demand and the capacity it can sustainably provide.

1

Demand is rising

Existing communities are scaling up; new STFC and wider DRI activities add demand.

2

Capacity is declining

Large earlier purchases are phasing out and replacement capital is insufficient.

3

Provision costs more

Hardware prices, running costs, maintenance and provider effort are all increasing.

4

£25m is not realistic

IRIS cannot expect capital alone to meet the full stated requirement.

5

How to close the gap

Lower cost, higher utilisation, efficient code, external capacity and tighter scrutiny.

6

Prioritisation is unavoidable

If scarcity persists, science-led choices must determine which activities receive more or less.

Sustain

Fund hardware and recurrent operation. Sweat the resources.

Broaden

Access NCR and other DRI capacity

Prioritise

Use evidence and authorised science governance

IRIS must move from buying hardware to managing a sustainable, mixed and explicitly prioritised capacity portfolio.

Discussion A: Obtain Additional Capacity

How can IRIS increase the sustainable capacity available to its communities?

1

Strengthen the capital case

Show the science consequences of the gap and the value delivered by additional investment.

2

Secure recurrent funding

Fund electricity, maintenance, warranties, provider staff effort and longer equipment lives.

3

Reduce purchase cost

Review memory, storage, node specifications and GPU flavours before procurement.

4

Access NCR and DRI resources

Agree allocation routes and identify workloads suitable for external infrastructure.

5

Use partner and spare capacity

Explore institutional, partner, national and genuinely unused DRI resources.

6

Build a mixed provision model

Integrate owned, partner and externally allocated capacity into a dependable service.

Capacity growth requires capital, recurrent resource, lower unit costs and credible access beyond IRIS-owned hardware.

Discussion B: Manage Persistent Scarcity

If additional capacity remains insufficient, how should IRIS reduce and distribute unmet demand?

1

Scrutinise demand more strongly

Challenge forecasts, specifications, growth assumptions and the timing of requests.

2

Monitor and adjust allocations

Track use, recover persistent under-use and revise allocations as evidence changes.

3

Improve effective utilisation

Raise occupancy and support more efficient code, workflows and data movement.

4

Choose the allocation principle

Continue broadly equal reductions, or protect selected activities and reduce others more.

5

Define science-led criteria

Consider commitments, strategic priority, consequences of reduction and alternative provision.

6

Authorise the decision body

Clarify whether CAP, STFC science boards or a dedicated cross-community panel should decide.

RSAP scrutinises need; IRIS models options; an authorised science body makes cross-community priority choices.

Discuss