



AI IN EDINBURGH AND THE UK

Professor Mark Parsons

EPCC Director
Dean of Research Computing

| epcc |

Summary

- UK Exascale preparations
- A regional data facility supporting AI
- The UK's AI Research Resource programme



Computer Room 4

£20m – CR 4 + PR D

£9.6m – 30MVA additional power

Space for 270 standard racks

Opened Dec 2020



ACF
confirmed
as Exascale
hosting site
in 2023

**PROPOSED POWER
GROUND FLOOR**

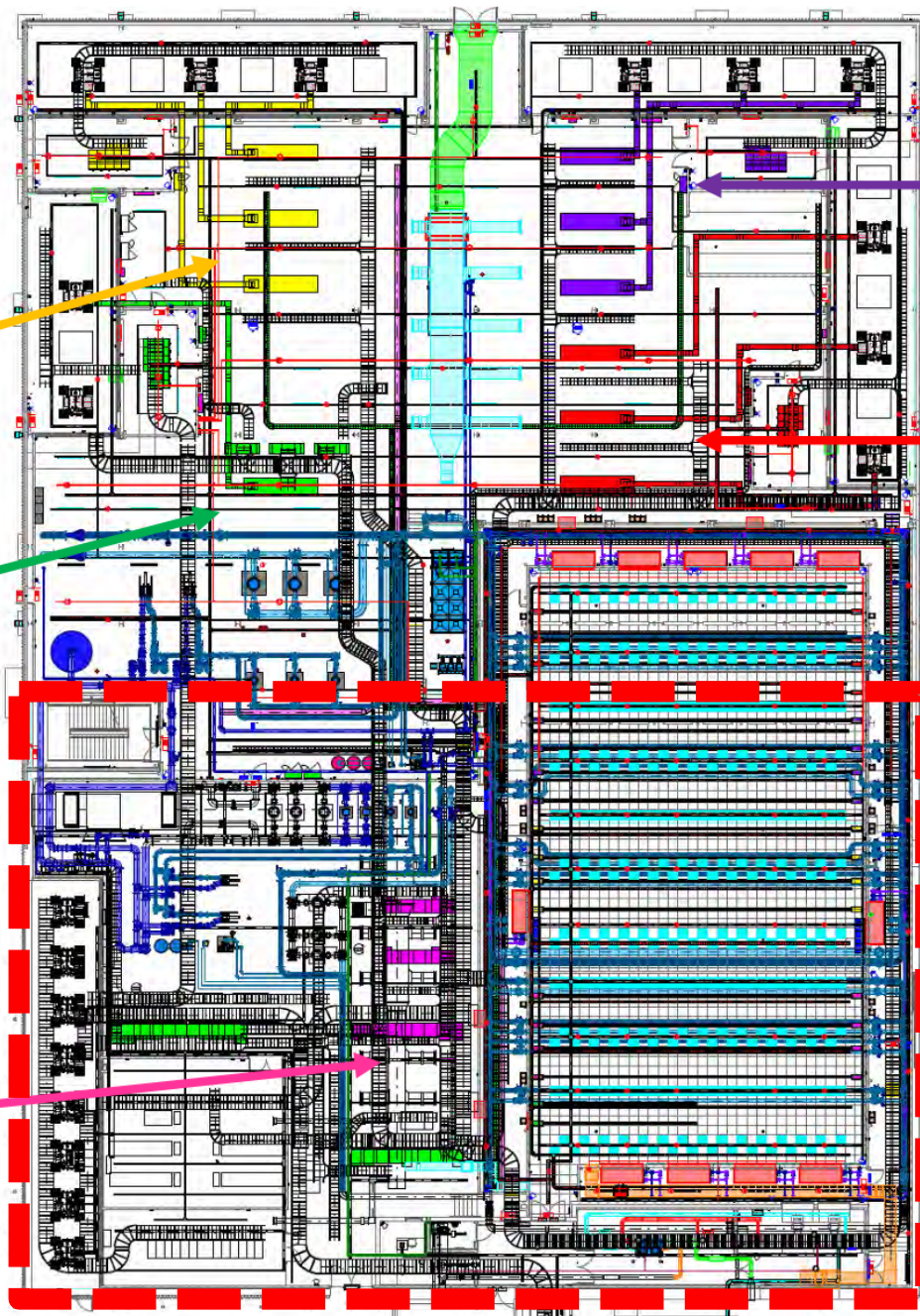
YELLOW STRING:
STRING 2 POWER FOR PHASE 2 HALL

GREEN STRING:
STRING 5 POWER FOR BUILDING
SERVICES/ UPS POWER.

PINK STRING:
STRING 1 POWER FOR PHASE 1 HALL

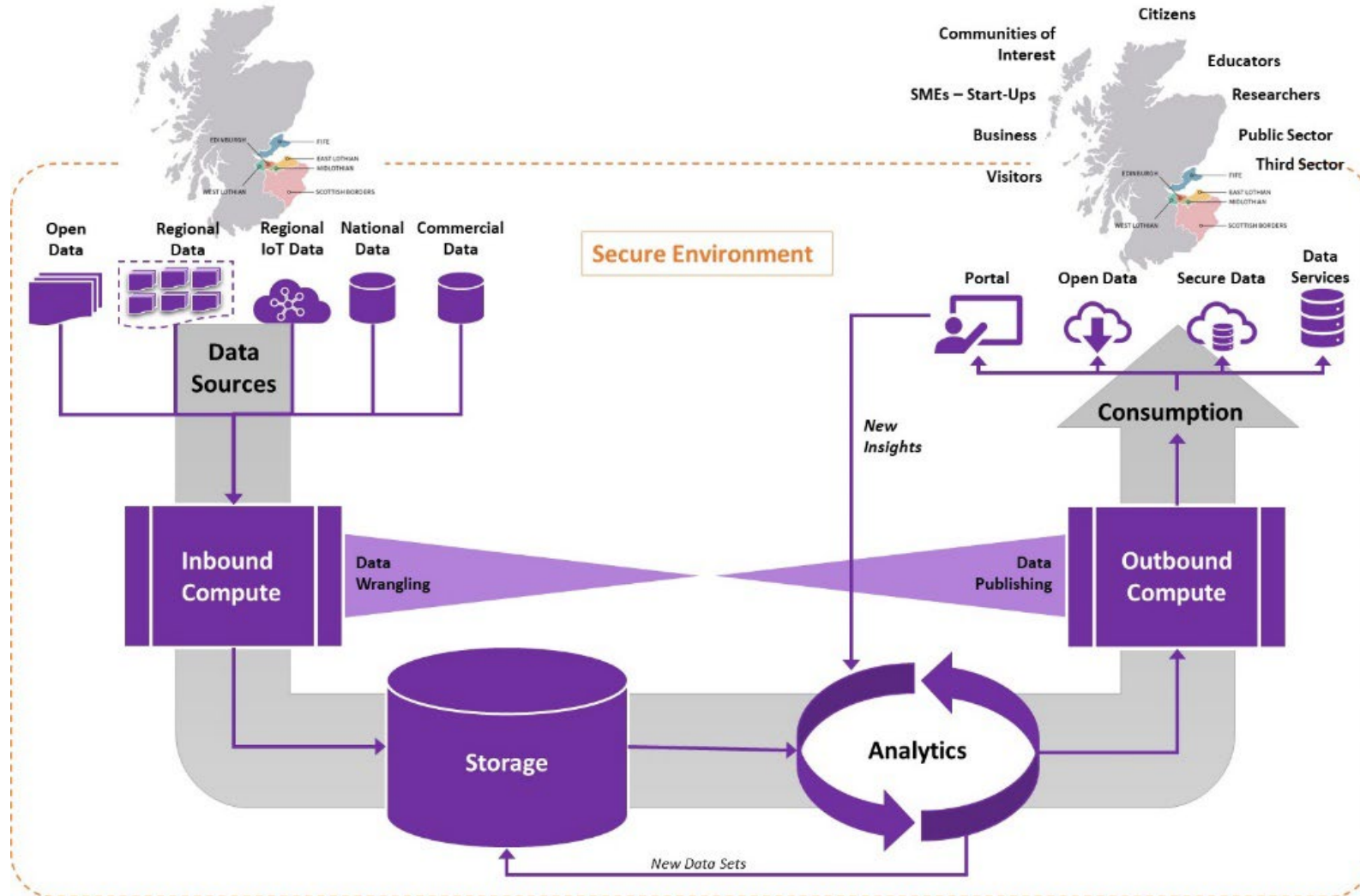
PURPLE STRING:
STRING 3 POWER FOR PHASE 2 HALL

RED STRING:
STRING 4 POWER FOR PHASE 2 HALL



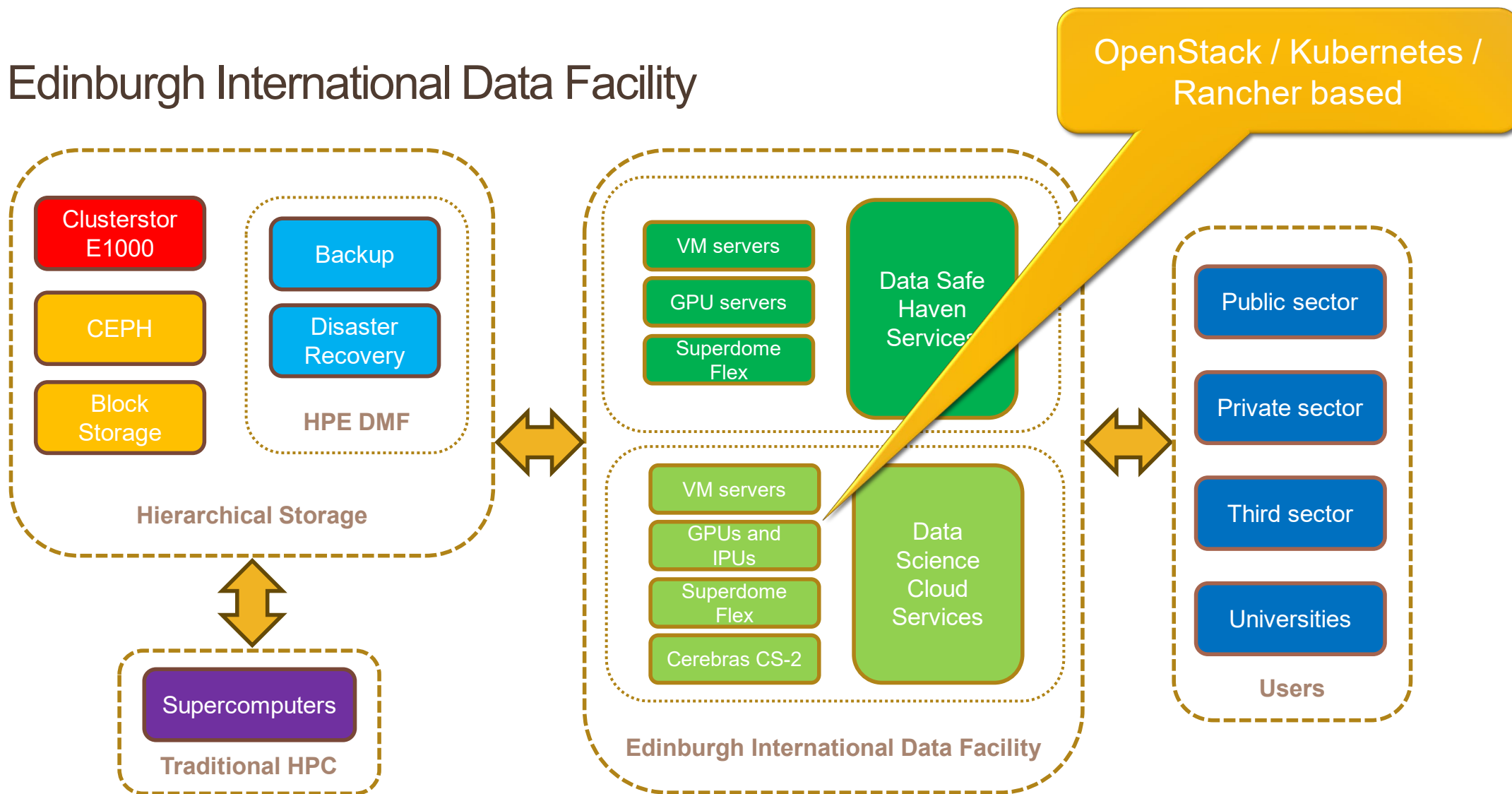
Total
investment
£93m

Designed for Data Science and AI projects and partnerships



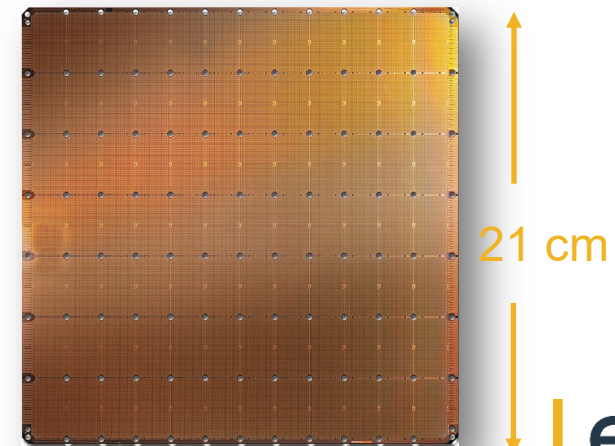
- Create a trusted **public-private-third sector** partnership
- Unlock economic opportunities worth **£5 billion+**
- Train **100,000 people** in data technologies

Edinburgh International Data Facility

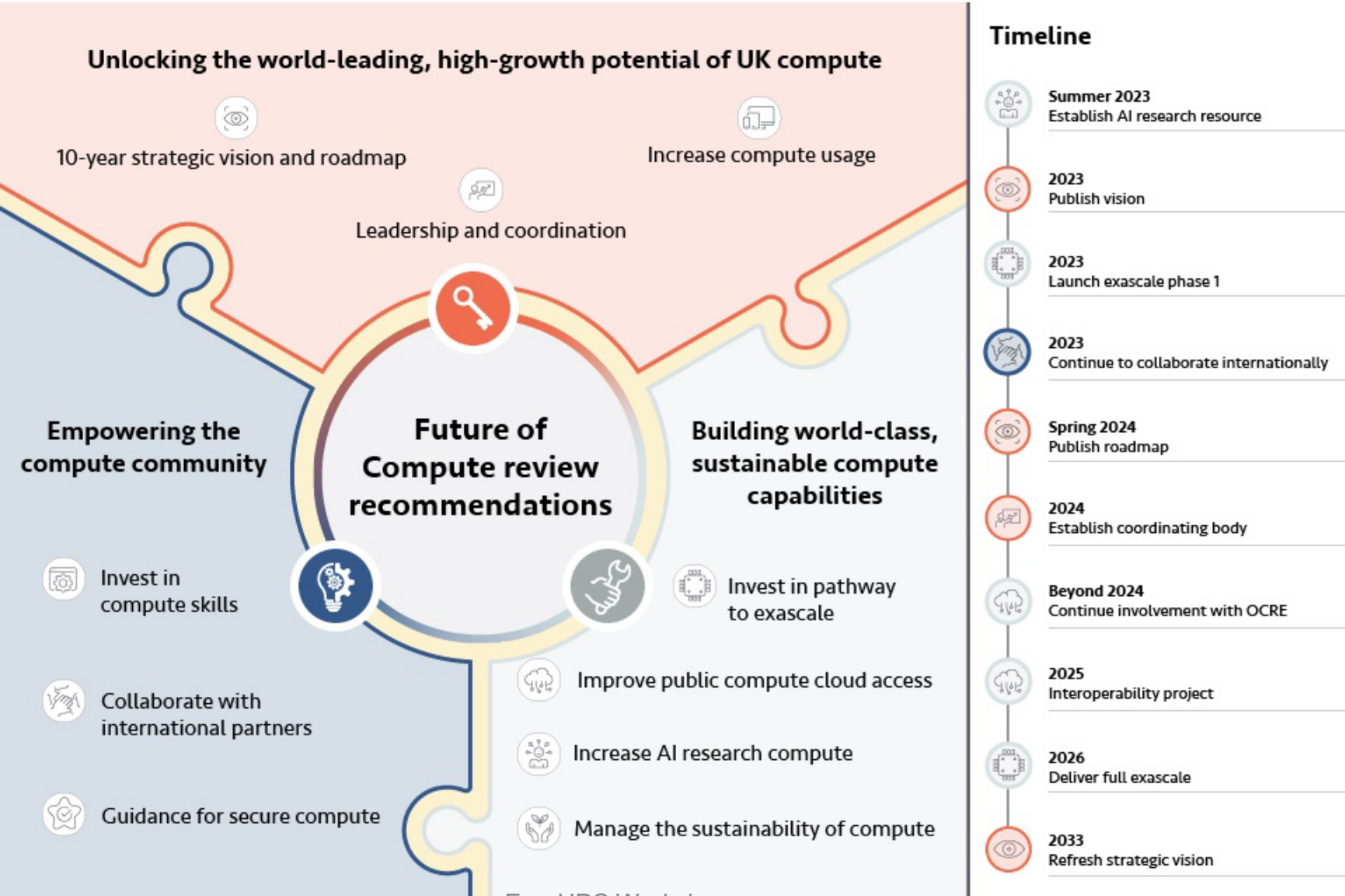


Beyond GPUs for AI ...

- EPCC installed Europe's first Cerebras CS-1 in March 2021
- Recently upgraded to two clustered CS-2s
- Tensorflow and PyTorch supported
- Focus on LLMs and large scale training
- User base gradually growing – difficult to move users from Nvidia GPUs ...



Future of Compute Review



Government policy

- Future of Compute Review – independent review published in Spring 2023. Highlighted importance of compute to UK R&D, and UK's increasing loss of international competitiveness in compute.
- Spring Budget 2023 - Chancellor immediately accepted two review recommendations, committing £900m to Exascale supercomputer and the AI Research Resource (AIRR).
- November 2023, Government announced first phase of AIRR 'Isambard-AI' in Bristol and 'Dawn' in Cambridge.
- Autumn Statement 2023, government announced further £500m to expand AIRR programme.

AIRR USE CASES

While AIRR will be open to a host of research domains and methods, some key use cases are likely to be:

- Large-scale training and fine-tuning of models like LLMs
- Training foundation models to tackle discrete scientific questions
- Wider AI safety research

CORE AIRR USERS

1 Fundamental AI research

2 Applied AI research

3 AI Startups

4 Government use cases

Users will access AIRR through a federated access model spanning all systems

AI within UK Digital Research Infrastructure

- Approach: drive AI adoption in research and innovation by:
 - Recognising that users are not homogenous: expectations vary across communities and levels of experience.
 - Providing a high-quality user experience.
 - Prioritising federation to reduce barriers across the UK's public digital infrastructure.
 - Investing in the development of Research Technical Professionals and supporting users in the transition to today's systems
 - Ensuring that compute access is secure and energy efficient.

