Federated Single Sign-On and Zero Trust Co-design for AI and HPC Digital Research Infrastructures

Sadaf R. Alam, Christopher Woods, **Matt Williams**, Dave Moore, Isaac Prior,
Ethan Williams, Fan Yang-Turner, Matt Pryor (StackHPC), Ilja Livenson (OpenNode)

University of BRISTOL

BriCS
Bristol Centre for Supercomputing

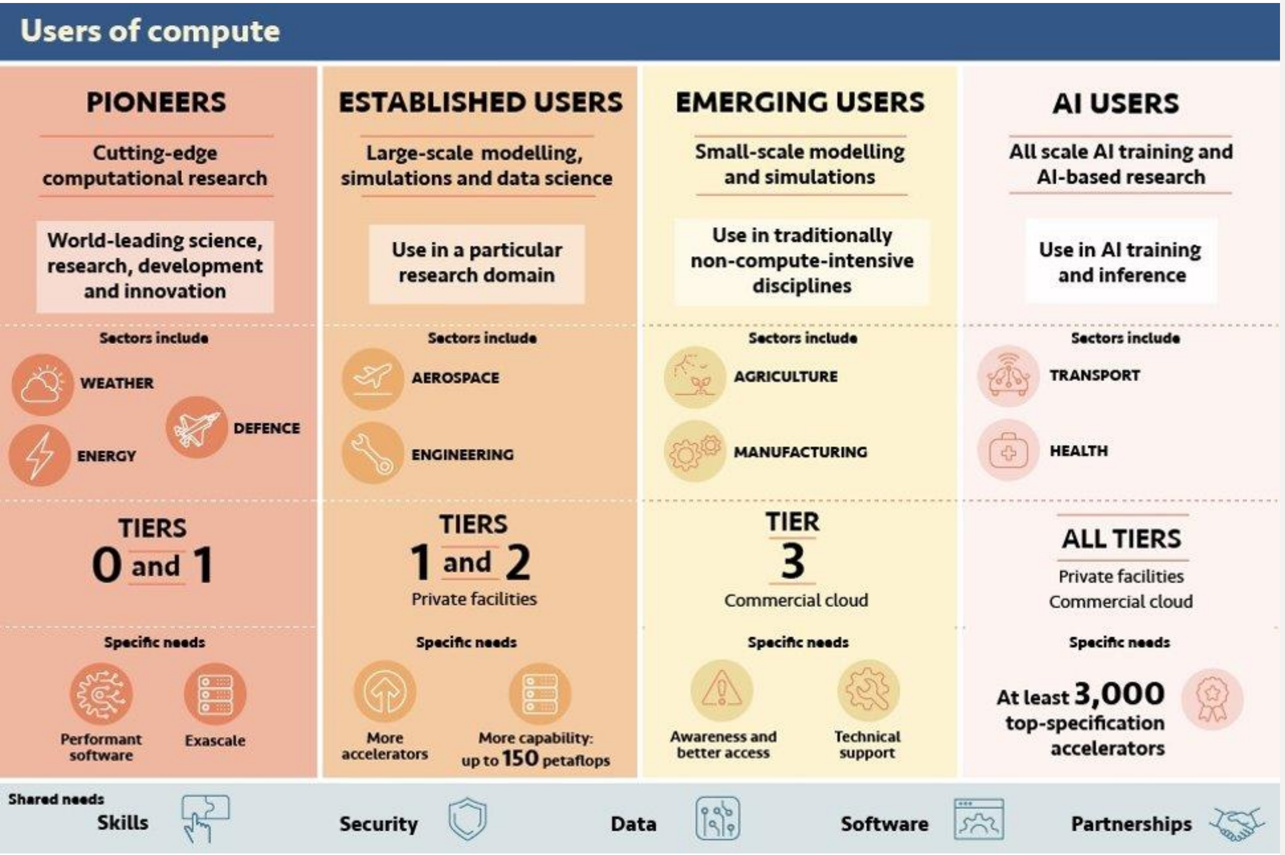
Isambard-AI

A UK national AI research resource

Independent report

Independent Review of The Future of Compute: Final report and recommendations

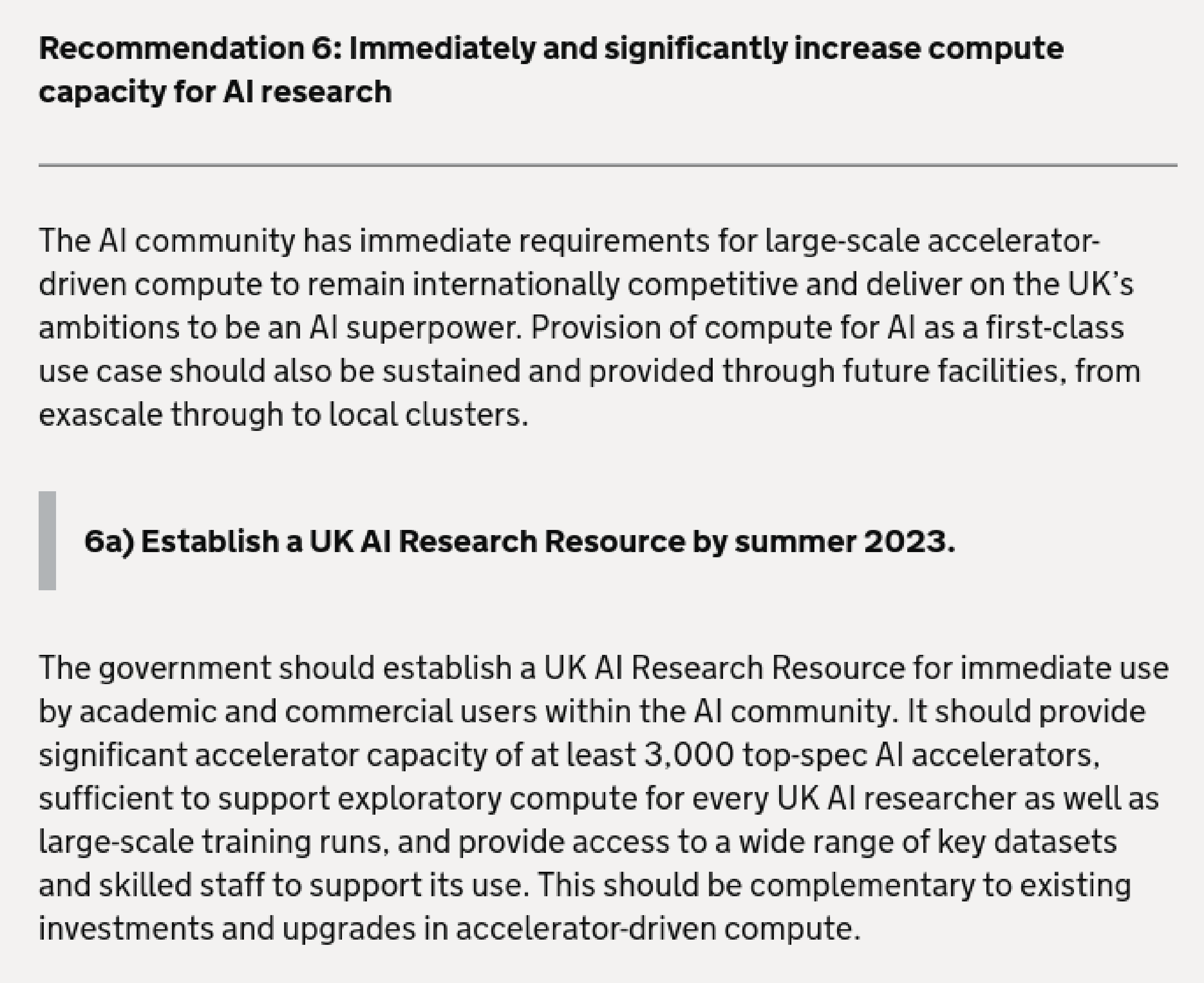
Updated 6 March 2023



Users of compute

PIONEERS	ESTABLISHED USERS	EMERGING USERS	AI USERS
Cutting-edge computational research	Large-scale modelling, simulations and data science	Small-scale modelling and simulations	All scale AI training and AI-based research
World-leading science, research, development and innovation	Use in a particular research domain	Use in traditionally non-compute-intensive disciplines	Use in AI training and inference
Sectors include WEATHER DEFENCE ENERGY	Sectors include AEROSPACE ENGINEERING	Sectors include AGRICULTURE MANUFACTURING	Sectors include TRANSPORT HEALTH
TIERS 0 and 1	TIERS 1 and 2 Private facilities	TIER 3 Commercial cloud	ALL TIERS Private facilities Commercial cloud
Specific needs Performant software Exascale	Specific needs More accelerators More capability: up to 150 petaflops	Specific needs Awareness and better access Technical support	Specific needs At least 3,000 top-specification accelerators

Shared needs: Skills, Security, Data, Software, Partnerships

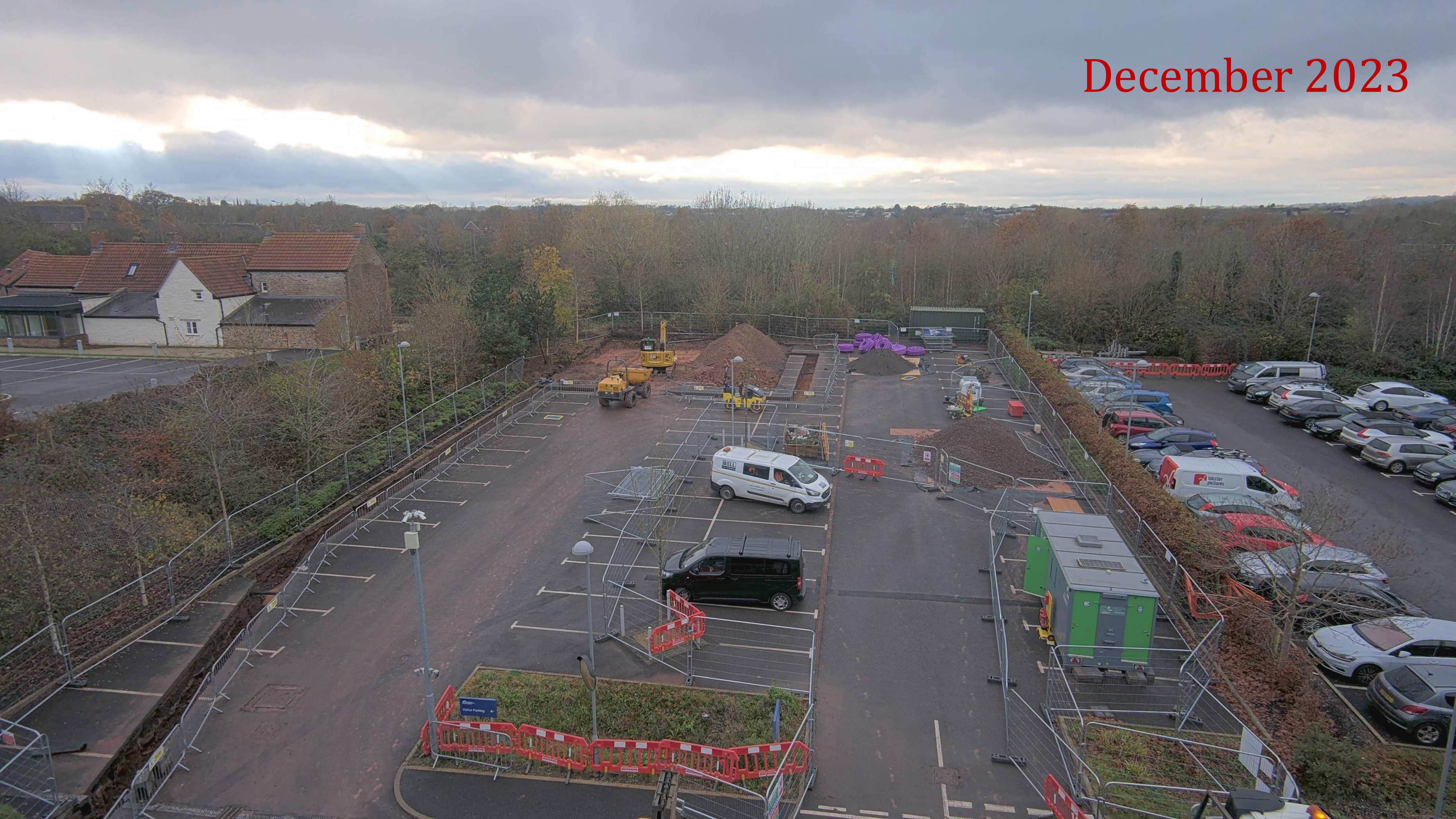


Recommendation 6: Immediately and significantly increase compute capacity for AI research

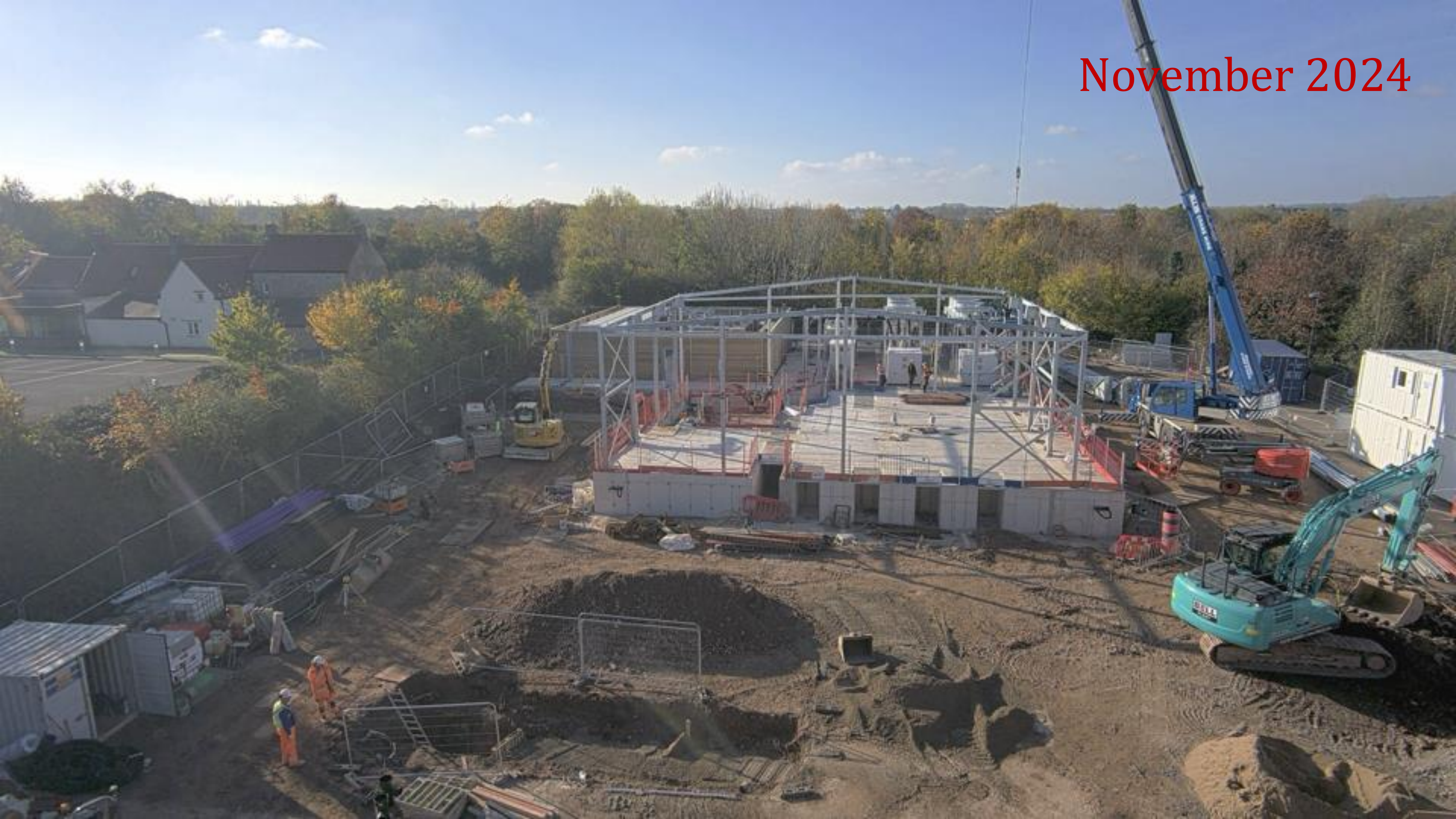
The AI community has immediate requirements for large-scale accelerator-driven compute to remain internationally competitive and deliver on the UK's ambitions to be an AI superpower. Provision of compute for AI as a first-class use case should also be sustained and provided through future facilities, from exascale through to local clusters.

> **6a) Establish a UK AI Research Resource by summer 2023.**

The government should establish a UK AI Research Resource for immediate use by academic and commercial users within the AI community. It should provide significant accelerator capacity of at least 3,000 top-spec AI accelerators, sufficient to support exploratory compute for every UK AI researcher as well as large-scale training runs, and provide access to a wide range of key datasets and skilled staff to support its use. This should be complementary to existing investments and upgrades in accelerator-driven compute.



December 2023



November 2024

Isambard 3 and Isambard-AI

Isambard-AI phase 1

- 42 nodes
- 168 Grace Hoppers
- 1 PB storage

Isambard 3

- 384 nodes
- 55,296 ARM cores

MDC 1

Isambard-AI phase 2

- 1,320 nodes
- 5,280 Grace Hoppers
- 27 PB storage

MDC 2

Foundations

Foundations

Zero-trust architecture

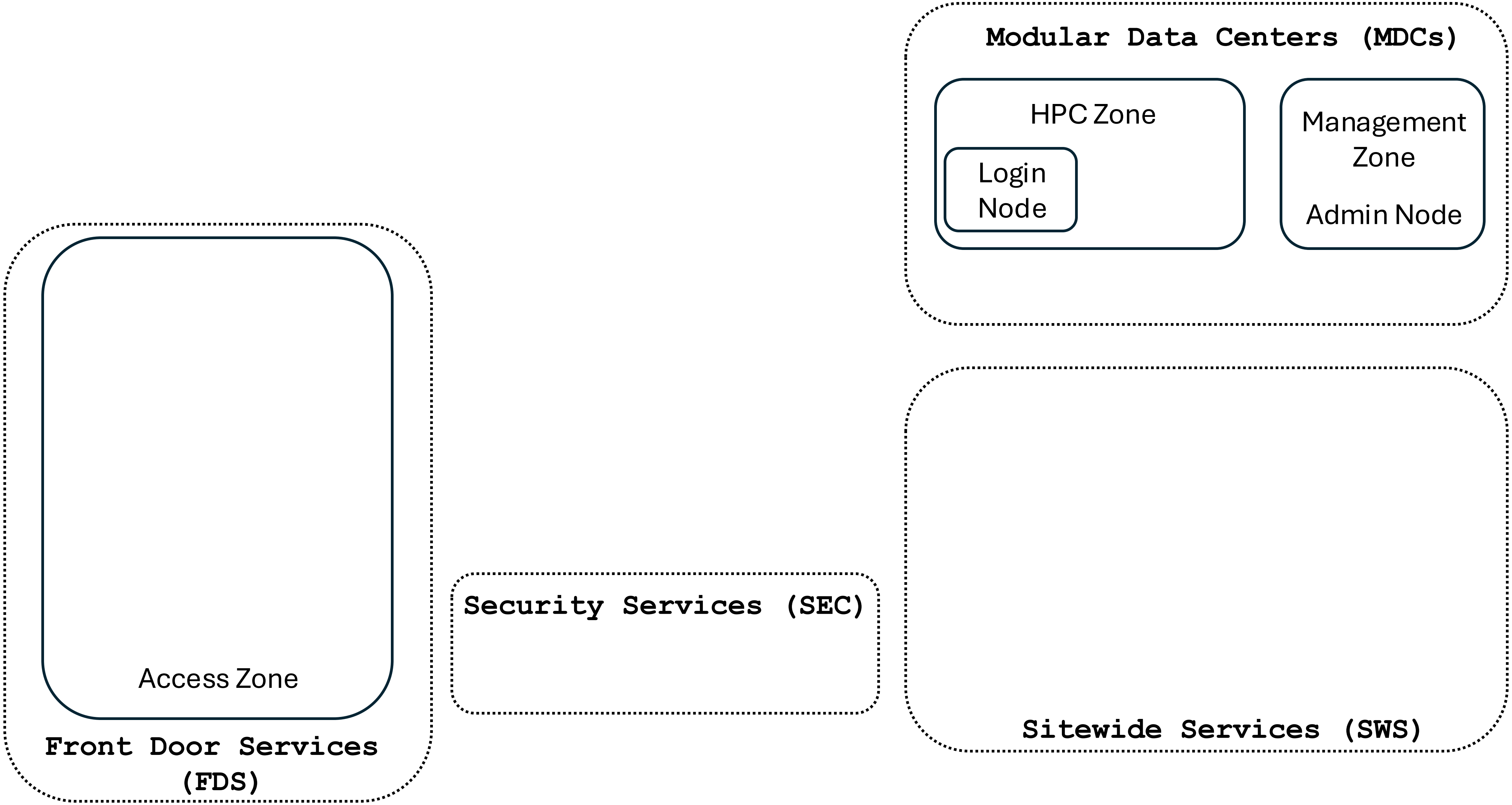
- Follows *NIST SP 800-223 [1]*
- Divide service into zones
- Extensive use of RBAC

Federated identities

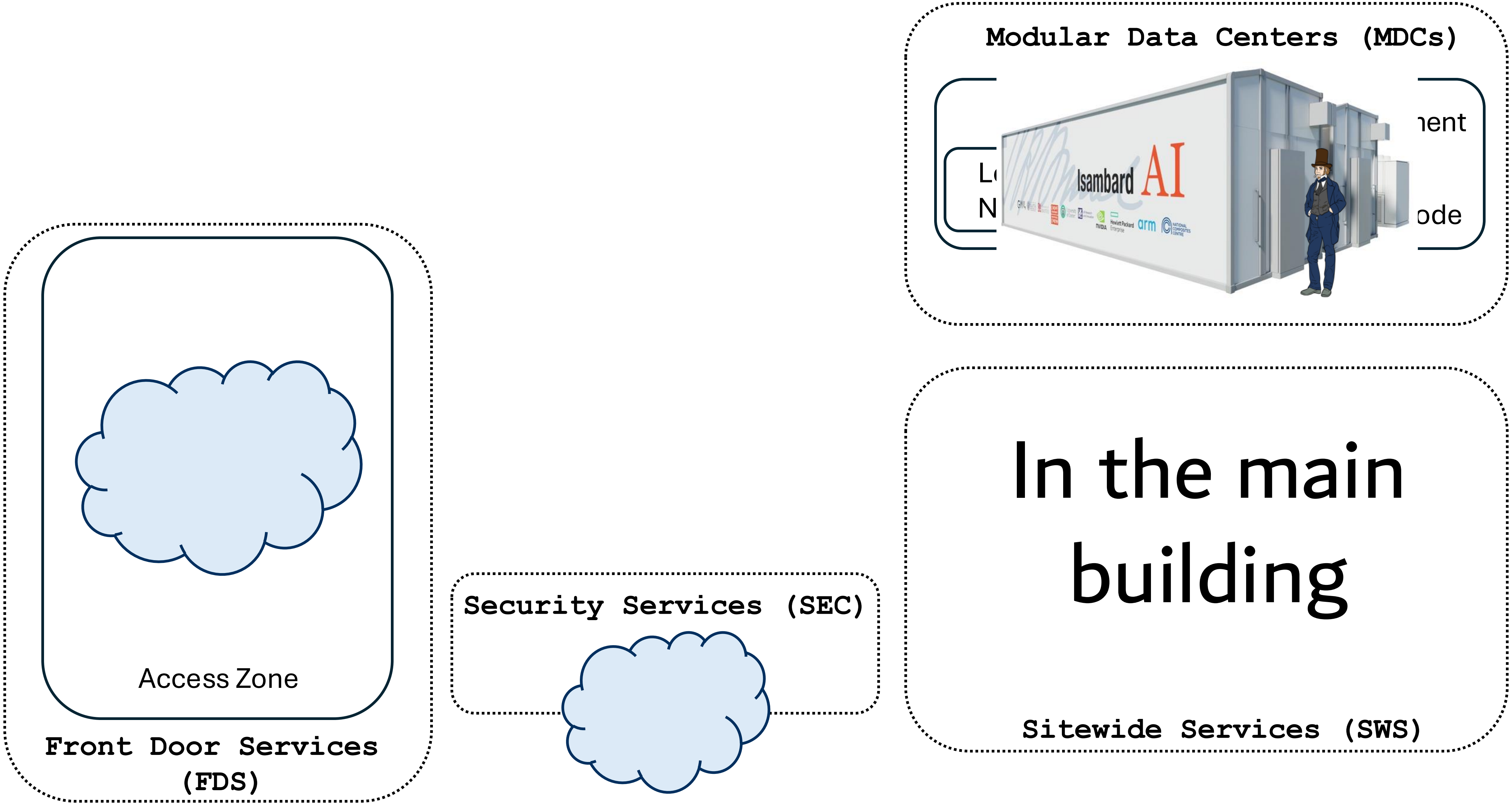
- Decentralise user identity management
- Separate *identity* from *access*

[1] 10.6028/NIST.SP.800-223

The zones


Modular Data Centers (MDCs)
HPC Zone
Login Node
Management Zone
Admin Node
Access Zone
Front Door Services (FDS)
Security Services (SEC)
Sitewide Services (SWS)

Where physically?



Modular Data Centers (MDCs)

L

N

nent

ode

In the main building

Sitewide Services (SWS)

Security Services (SEC)

Access Zone

Front Door Services (FDS)

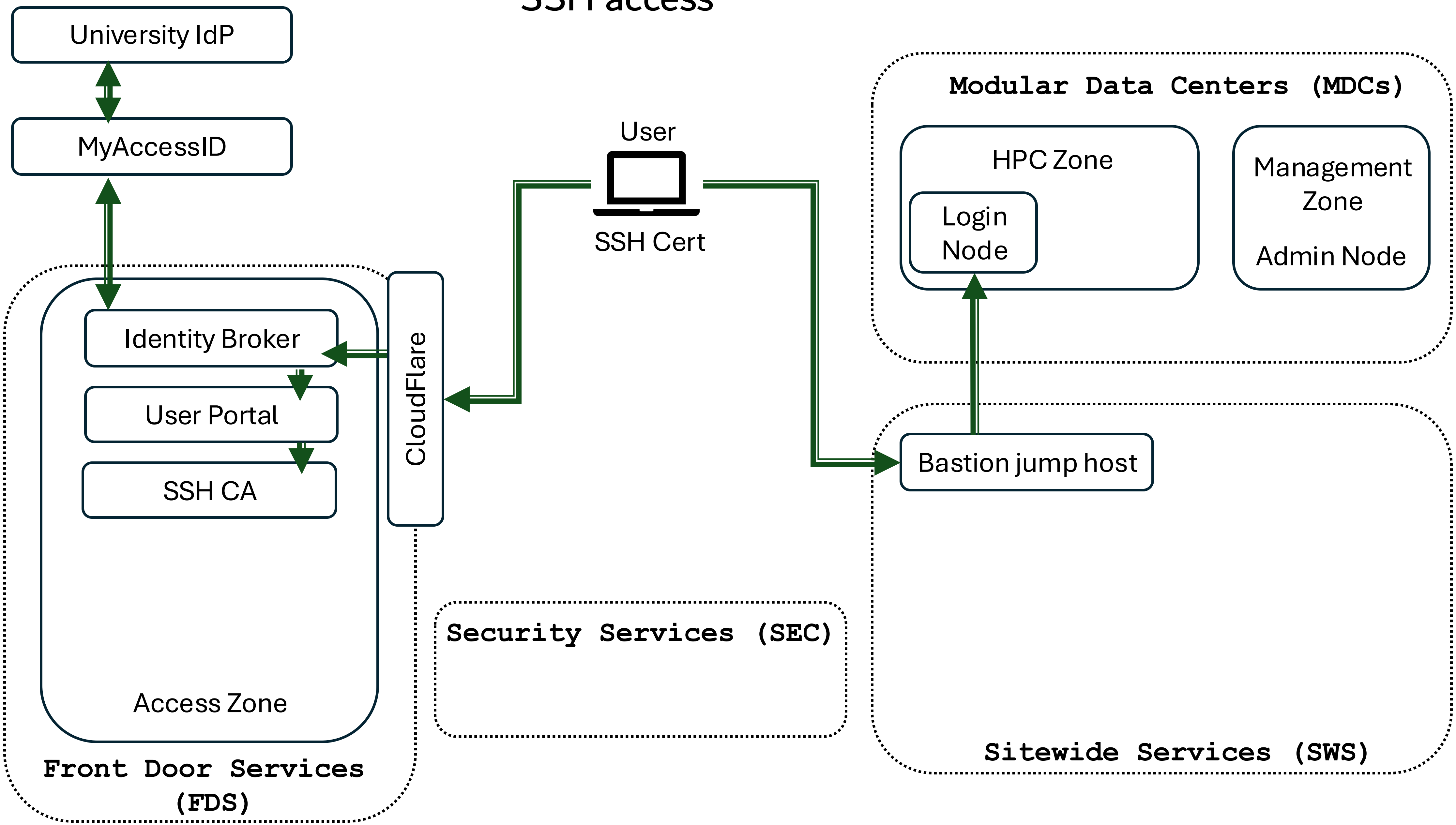
User access

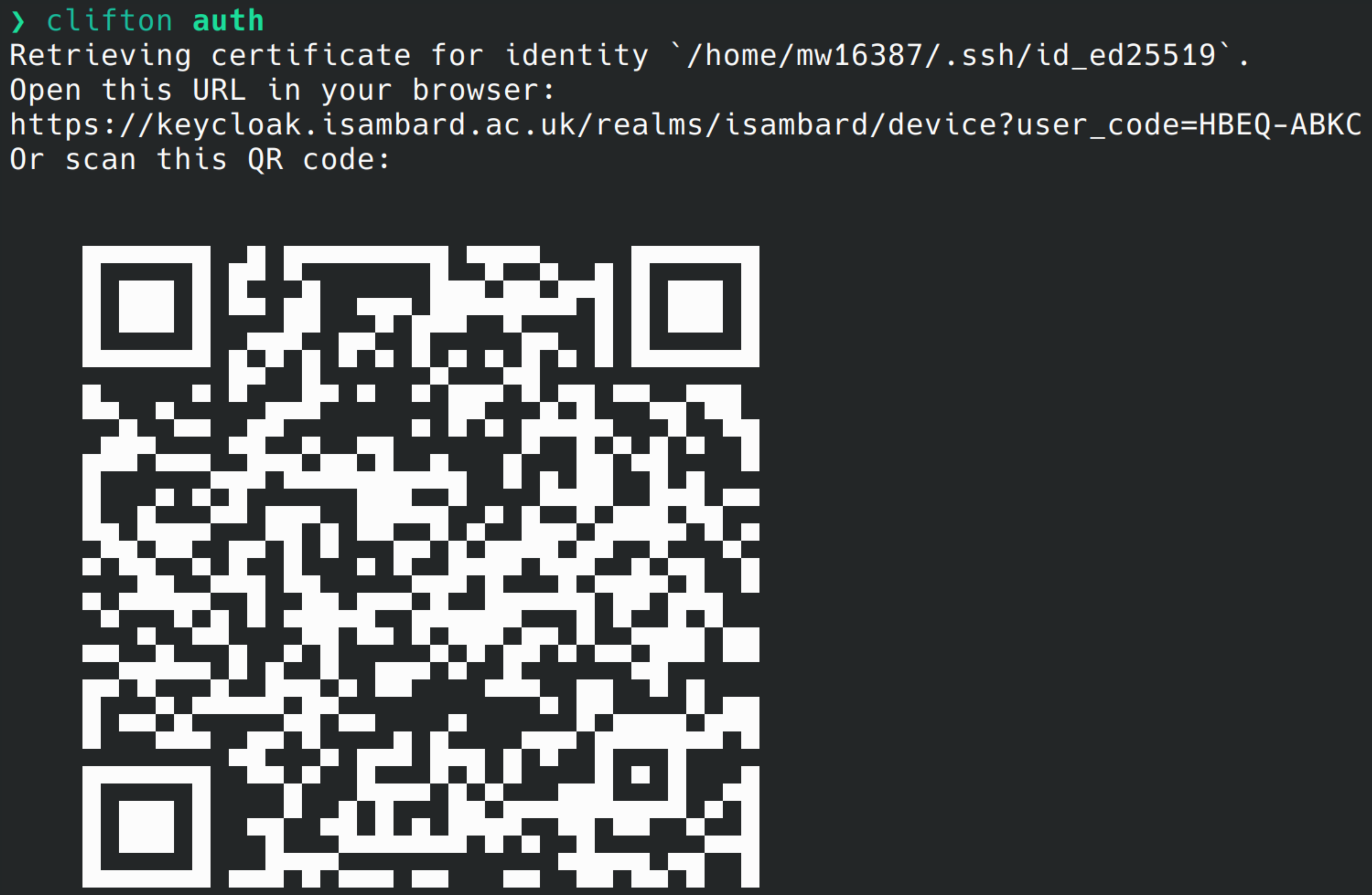
Federated identities

- Traditional HPC centres create accounts for users of their system – including managing passwords and/or SSH keys.
- We don't want to hold credentials for users.
- We broker out to MyAccessID to authenticate users.

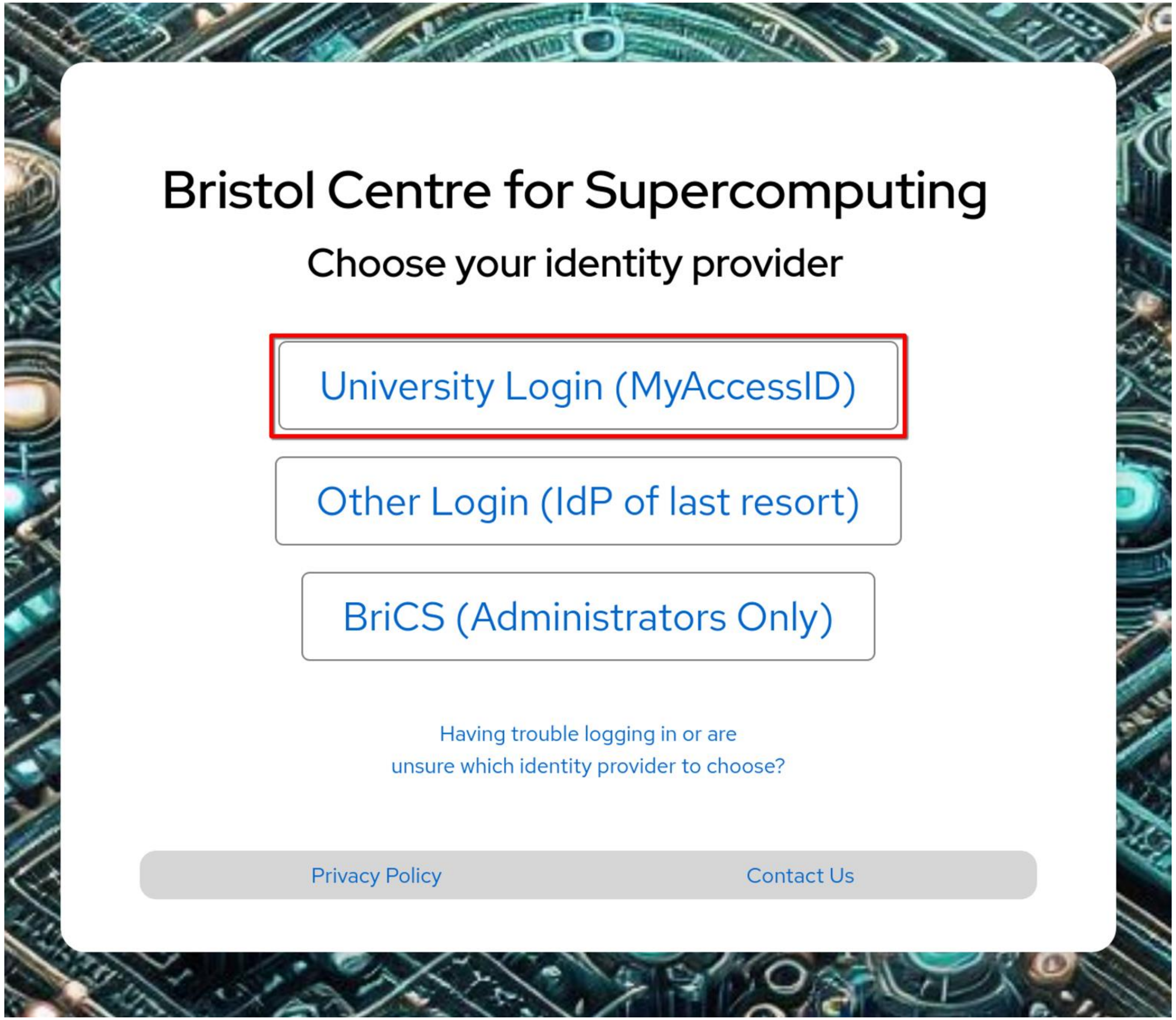
- Given an authenticated identity, we hold access lists to control who can access which resources.

SSH access


University IdP
MyAccessID
User
SSH Cert
Modular Data Centers (MDCs)
HPC Zone
Login Node
Management Zone
Admin Node
Identity Broker
User Portal
SSH CA
CloudFlare
Access Zone
Front Door Services (FDS)
Security Services (SEC)
Bastion jump host
Sitewide Services (SWS)



```
❯ clifton auth
Retrieving certificate for identity `/home/mw16387/.ssh/id_ed25519`.
Open this URL in your browser:
https://keycloak.isambard.ac.uk/realms/isambard/device?user_code=HBEQ-ABKC
Or scan this QR code:
```



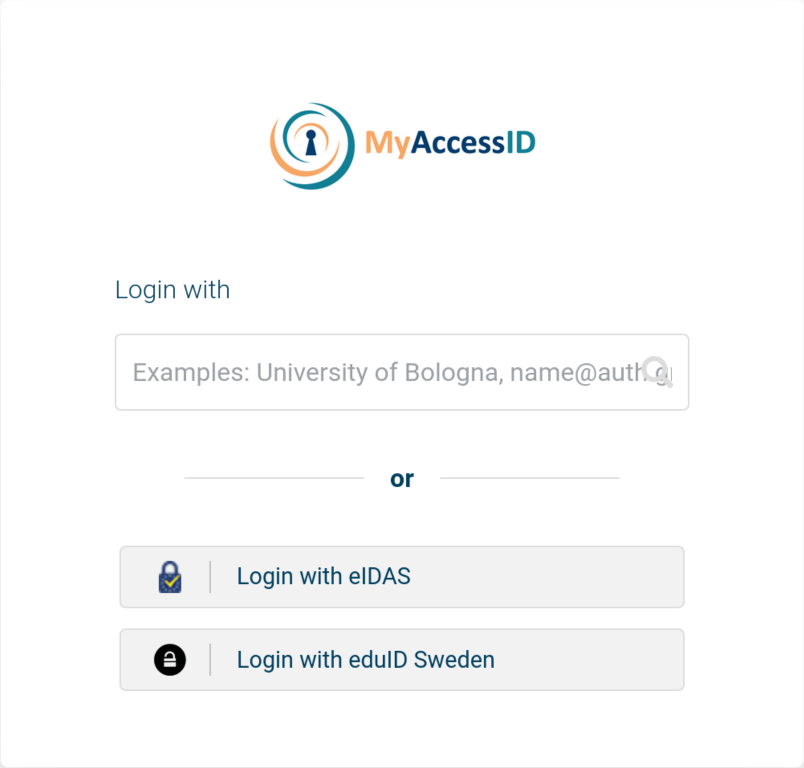
Bristol Centre for Supercomputing

Choose your identity provider

University Login (MyAccessID)
Other Login (IdP of last resort)
BriCS (Administrators Only)

Having trouble logging in or are
unsure which identity provider to choose?

Privacy Policy
Contact Us



MyAccessID

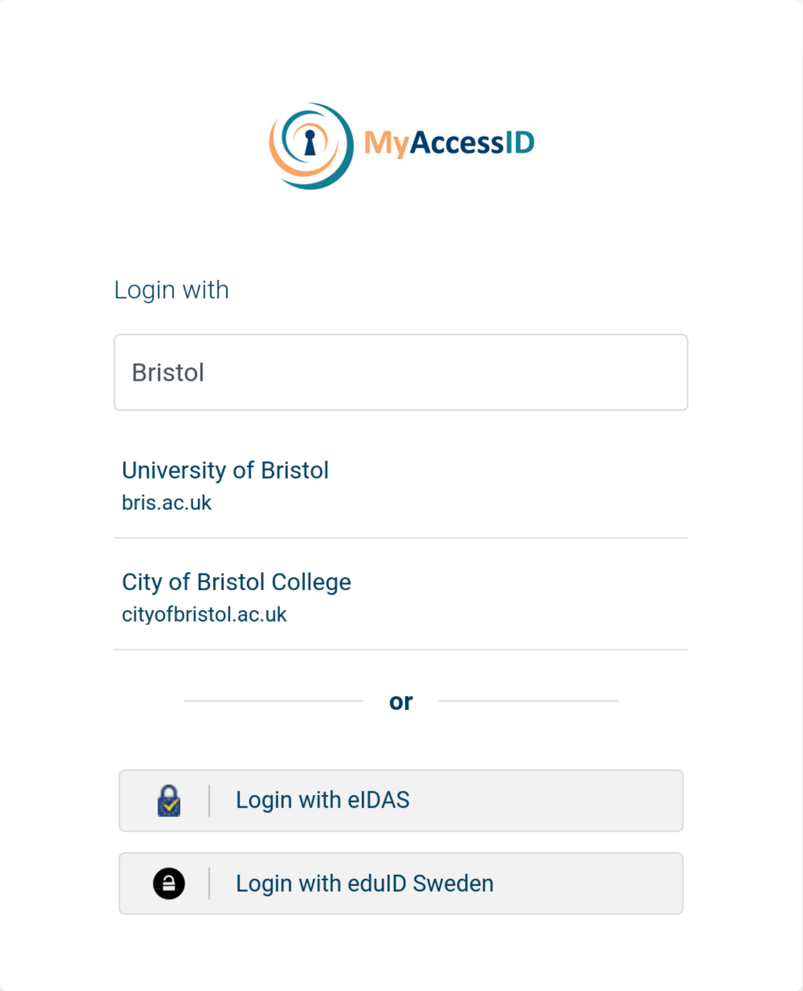
Login with

Examples: University of Bologna, name@auth

or

Login with eIDAS

Login with eduID Sweden

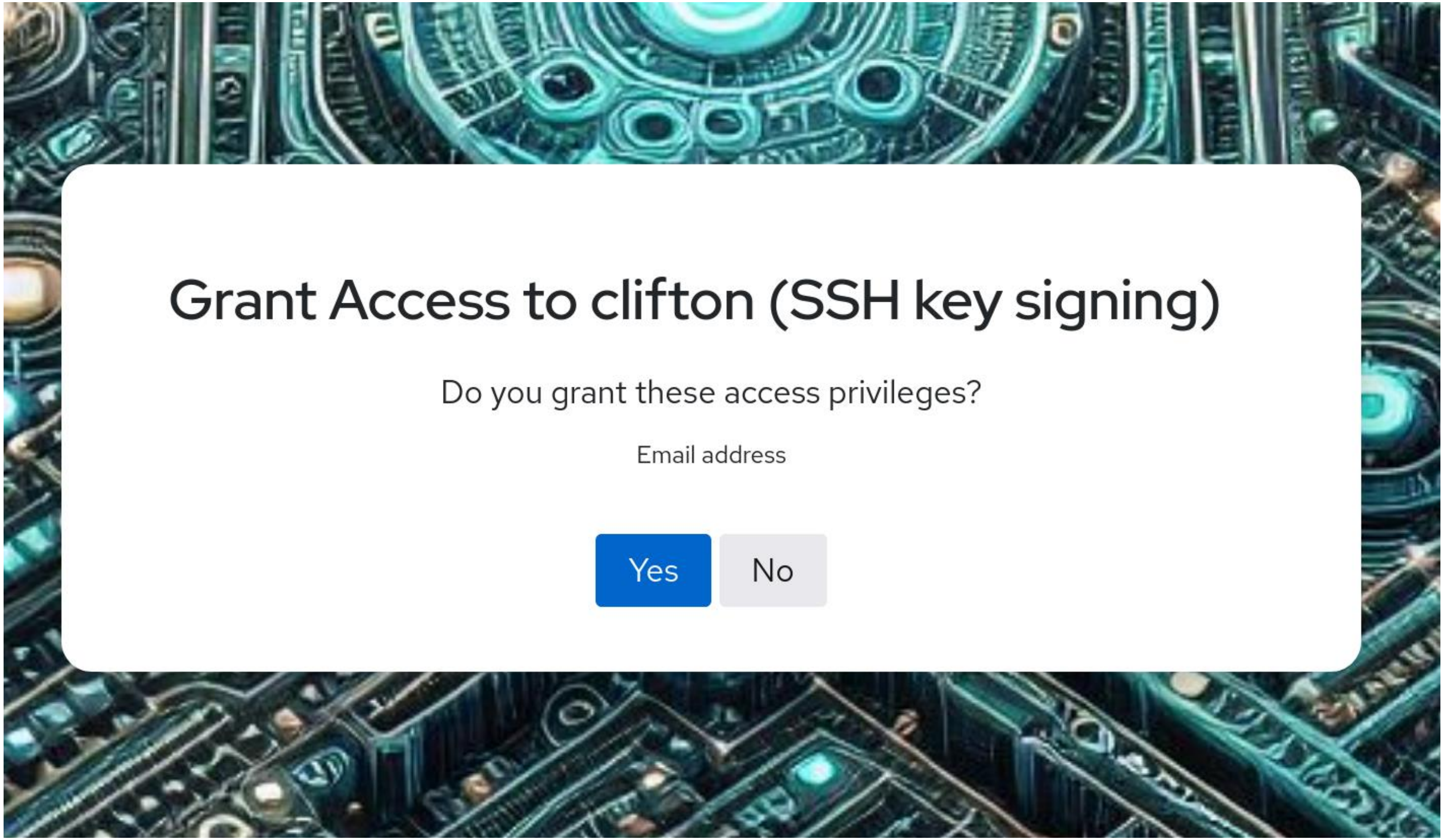


Login with

Bristol

University of Bristol
bris.ac.uk

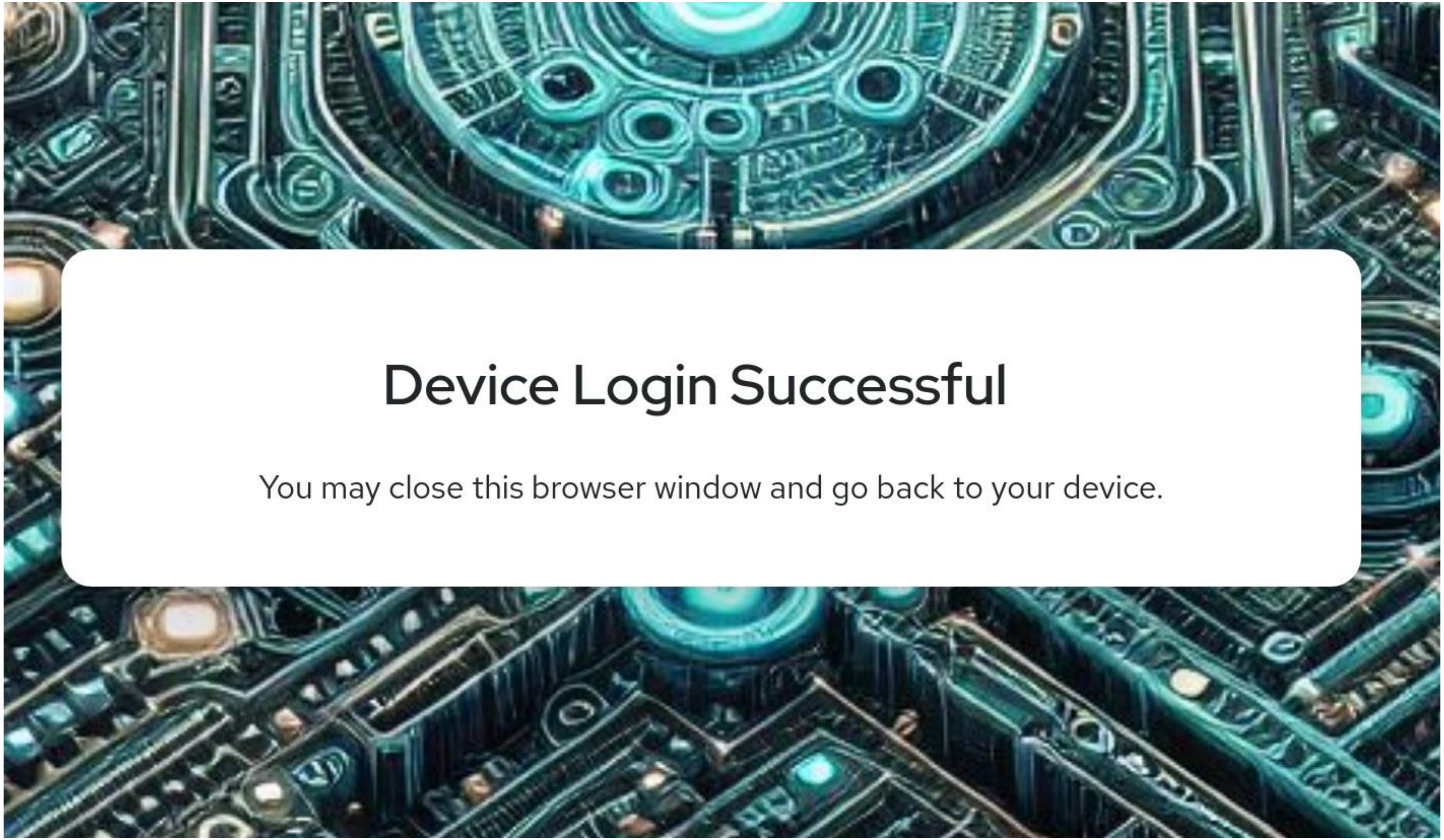
City of Bristol College
cityofbristol.ac.uk



Grant Access to clifton (SSH key signing)

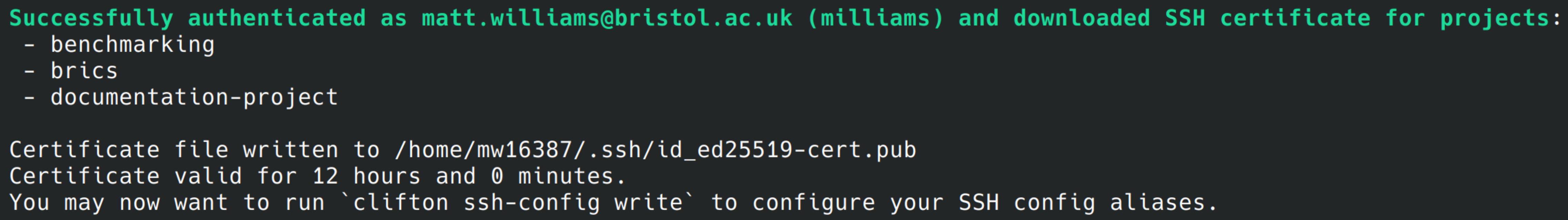
Do you grant these access privileges?

Email address



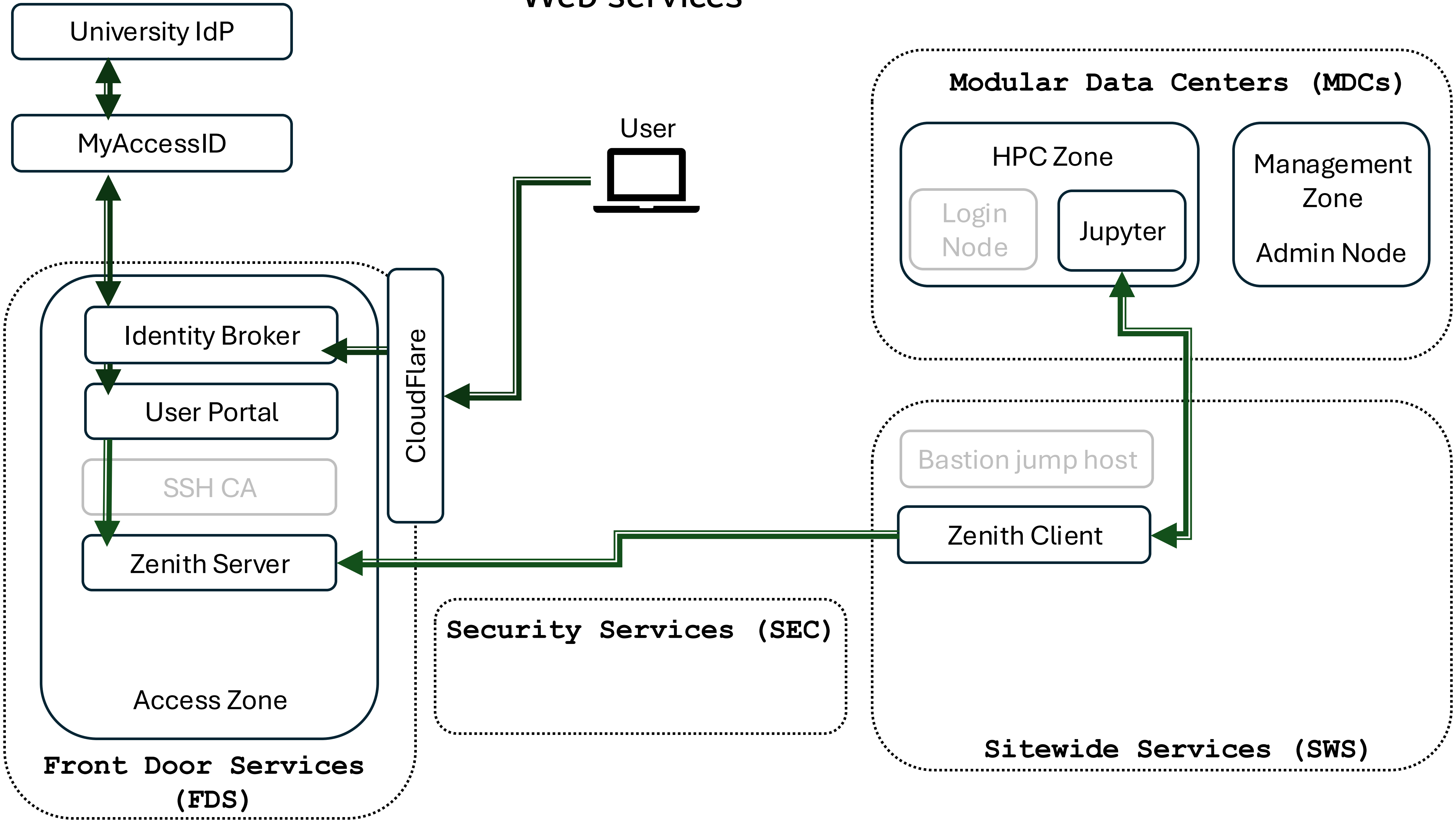
Device Login Successful

You may close this browser window and go back to your device.



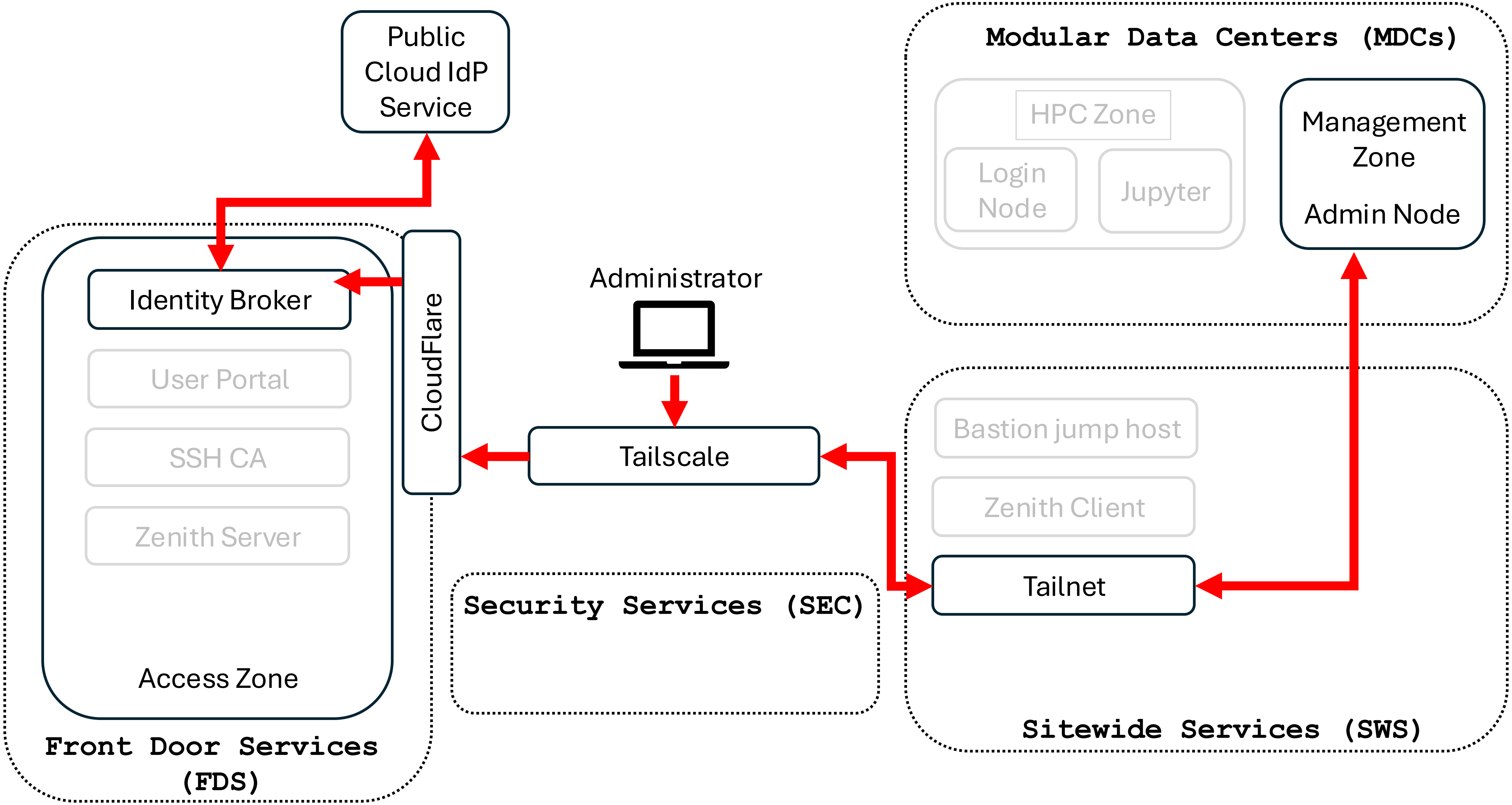
```
Successfully authenticated as matt.williams@bristol.ac.uk (milliams) and downloaded SSH certificate for projects:
 - benchmarking
 - brics
 - documentation-project

Certificate file written to /home/mw16387/.ssh/id_ed25519-cert.pub
Certificate valid for 12 hours and 0 minutes.
You may now want to run `clifton ssh-config write` to configure your SSH config aliases.
```

Web services

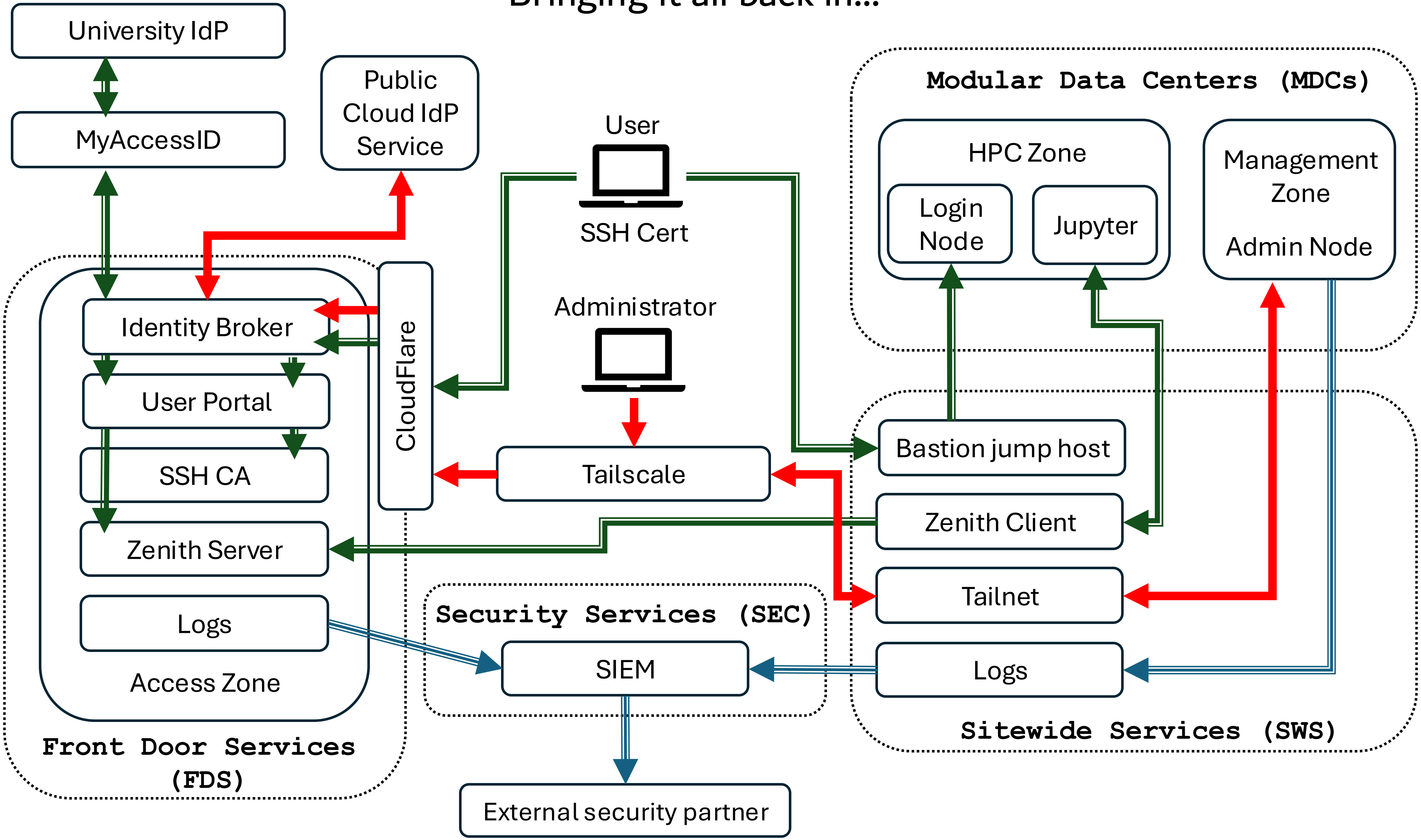


Admin access

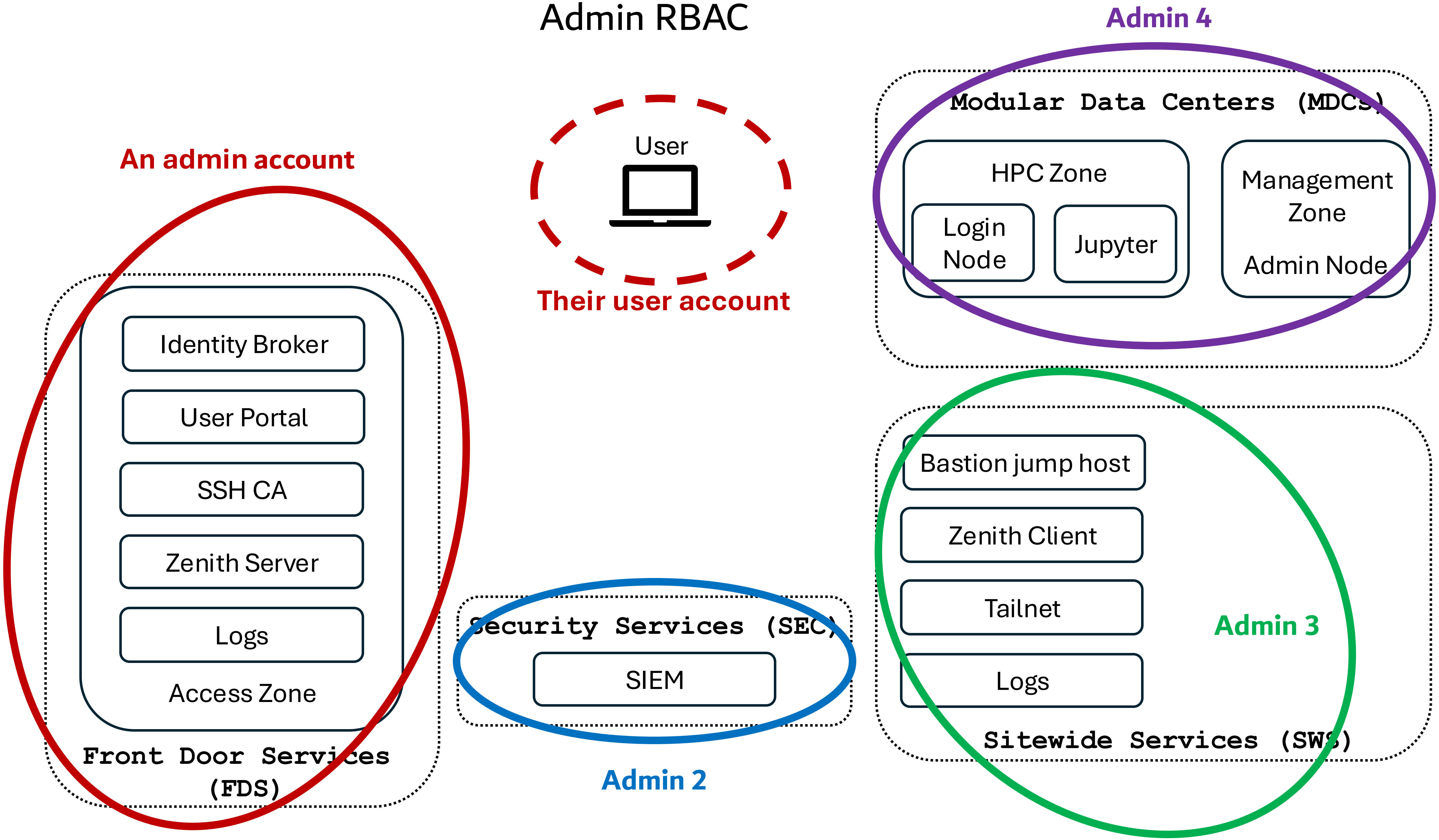
Admin access



Bringing it all back in...



Admin RBAC



Admin 4

Modular Data Centers (MDCs)

HPC Zone

Login Node

Jupyter

Management Zone

Admin Node

An admin account

User

Their user account

Identity Broker

User Portal

SSH CA

Zenith Server

Logs

Access Zone

Front Door Services (FDS)

Security Services (SEC)

SIEM

Admin 2

Bastion jump host

Zenith Client

Tailnet

Logs

Admin 3

Sitewide Services (SWS)

Summary

Zone-based architecture

- Users and admins are kept separate
- Users access is controlled in the access zone
- Admin access is specific to their role
- Service→service is zero-trust

Federated user identities

- User identities come from their institution
- Identities are resolved consistently before access control
- Access to resources is granted on a project-basis

[1] 10.6028/NIST.SP.800-223

Thank you

Dr Matt Williams
matt.williams@bristol.ac.uk

brics-enquiries@bristol.ac.uk

University of BRISTOL

BriCS
Bristol Centre for Supercomputing