

Our Future Health

Cloud Infrastructure

Procurement – Requirements

Version: 03 March 2022

Contents

1	Public cloud services.....	1
1.1	Basic services	1
1.2	APIs and tooling	2
1.3	Data storage services	2
1.4	Data ingress.....	2
1.5	Data workflows.....	2
1.6	Data encryption.....	2
2	Data centres and location.....	3
2.1	Hosting location	3
2.2	Backups	3
2.3	Recovery point	3
2.4	Recovery time.....	3
3	Computation	3
4	Hosting of a Trusted Research Environment (TRE).....	4
4.1	Availability of TRE providers	4
4.2	Researcher billing.....	4
4.3	Pricing for researchers	4
4.4	Vouchers for researchers	4
5	Future-proofing and portability	4
6	Technical assistance	5
7	Support.....	5

1 Public cloud services

1.1 Basic services

The cloud provider shall provide a comprehensive set of standard compute, storage, and infrastructure services. This list includes, but is not limited to, the kinds of core cloud services we would expect to see:

- Virtual machines
- Object storage
- Managed relational database
- Managed non-relational database
- Managed load balancers

- Monitoring and integrated autoscaling
- Serverless computing
- DevOps tools
- Backup and recovery
- Archiving
- Content delivery
- Visibility via dashboards
- Other connected and related products and services

1.2 APIs and tooling

We will use infrastructure-as-code (IaC) and associated best practices to create and manage cloud services. The cloud shall provide APIs which allow programmatic management of all services and shall have API wrapper libraries in all popular programming languages and good support in popular IaC tools such as Terraform and Ansible. The cloud shall have a community of users using these libraries and tools.

1.3 Data storage services

Our Future Health will store a large amount of data with variable availability requirements (see “Data storage” section below for more details). In terms of data storage services provided by the cloud, it shall have a tiered structure where cost of storage is traded off against the availability and retention period of the data.

1.4 Data ingress

Our Future Health will be receiving data from a variety of third-party sources (e.g. the NHS, our bioinformatics suppliers). The cloud shall provide secure data transfer services which can land data in our cloud storage, including, but not limited to, things like API gateways and managed SFTP servers.

1.5 Data workflows

Our Future Health expects to build multiple data ingestion/ETL pipelines to bring in data from third-party sources. We expect to use a mixture of services provided by the cloud natively and third-party software and tools built on top of those services.

1.6 Data encryption

The cloud shall provide data storage solutions that allow for data to be encrypted while at rest using strong, current cryptographic algorithms (as recommended by the National Cyber Security Centre (NCSC)). The cloud shall allow us to manage the encryption keys such that the cloud provider does not have access to the decrypted data or provide strong controls to prevent such access.

2 Data centres and location

2.1 Hosting location

The cloud shall provide hosting of all required services within the UK.

2.2 Backups

Our primary data store represents the main value of our programme and so requires strong protection from loss or corruption, particularly the data where we are the source and so cannot be recovered from other partners. The cloud shall have the ability to have a backup that is physically and logically isolated and failure-independent from the main data storage location, even if it sits within the same geographic region.

In addition, we will face the risk of ransomware or other attacks, including on backup data. The cloud shall have the capability to store a backup that is considered sufficiently “offline” to defend against such an attack, according to the guidance such as that from NCSC’s “Offline backups in an online world”.

2.3 Recovery point

Our Recovery Point Objective is 24 hours for the primary data store, including from the offline backup described above. There will be additional, smaller data stores that have shorter RPOs, such as user registration data.

2.4 Recovery time

We can tolerate a loss of availability far more readily than a loss of integrity or confidentiality. Our Recovery Time Objective can be up to 2 calendar days.

3 Computation

We expect that we will use a variety of compute resources in our environment to power things like (but not limited to):

- Interactive web and app backend services
- Batch data processing and ETL jobs dealing with import, export and sharing of data, that may include bioinformatics jobs working with genetics data
- Batch data transfer jobs between us and our partners such as labs dealing with biospecimens, via APIs, SFTP or similar
- Monitoring, analytics, and business intelligence dashboards

As well as running our own compute workloads, there will also be large compute workloads generated by researchers inside the TRE. These will include tasks such as data analysis, statistical computing, machine learning and bioinformatics. It is not possible for us to predict the expected researcher workloads, but we note that they could easily exceed our own workloads.

4 Hosting of a Trusted Research Environment (TRE)

4.1 Availability of TRE providers

We are interested in working with cloud providers that have good availability of TRE providers already working within their systems, and we would like to see a list of available TRE providers. We will also be asking TRE providers who are bidding as part of our procurement which cloud providers they can run on within the specified launch time frame.

4.2 Researcher billing

As a separate lot within this procurement, we will be looking for a Supplier to handle the billing of researchers for their cloud usage within the selected TRE running on the selected cloud. The Researcher Billing Lot may be awarded to the TRE Supplier. In that instance, we will require our cloud Supplier to integrate with the selected billing Supplier and TRE Supplier such that:

- Our cloud Supplier (for Lot 1) must enable the billing Supplier to track the cloud costs and usage of researchers in our TRE
- Our cloud Supplier (for Lot 1) must enable the billing Supplier to manage billing accounts for researchers, collect any monies owed and settle any cloud costs they incur

We require that any one-off or ongoing integration costs the cloud Supplier incurs to create this billing relationship are at no additional cost to Our Future Health.

4.3 Pricing for researchers

Researchers working in our TRE should receive a saving on their compute costs and storage costs. This should be based on a discount from the regular list pricing on use of compute instances and relevant storage. Suppliers should provide details of discounts available and propose how they will ensure researchers do not pay higher rates for their cloud usage than Our Future Health.

4.4 Vouchers for researchers

We expect that some researchers applying to Our Future Health will not have significant funding and may require assistance in the form of reduced cloud costs. We would like cloud Suppliers to offer vouchers to cover initial costs for all researchers, and potentially ongoing costs for select researchers who have less access to funding.

5 Future-proofing and portability

Our Future Health is intended to be a very long-term programme, over the lifetimes of the participants.

We therefore need to design technology solutions that will be viable over multiple decades. One could argue that modern cloud computing is around 15 years old. Who knows what another 15 years will bring? It is therefore incumbent on us to ensure that we

can continually migrate our systems as technology and services evolve. More prosaically, a subsequent procurement may result in a different supplier winning this contract.

We must be prepared to migrate all services, data and configurations to a new Supplier at minimal exit cost, in particular minimising any charges relating to movement of our large data stores.

6 Technical assistance

We are aware that our cloud Supplier (for Lot 1) will likely have access to a much broader and deeper array of expertise than we would ever have in our own organisation. Our Future Health will not achieve our aims unless we can work collaboratively across the ecosystem and rely on others' expertise more than our own. We'd like to work with a cloud Supplier who can offer technical assistance and solutions architecture advice as we scale at no additional cost to us. We also require the cloud Supplier to offer technical training to our staff to ensure they are familiar with the available tools and best practices.

7 Support

We would like access to a support function from our cloud Supplier in case of technical issues particularly with our production systems. We would like to understand what SLAs can be offered but would expect to see:

- 24x7x365 access to cloud support
- A sliding scale of response times, with a maximum of 1 hour in the case of production systems that are down, 365 days per year