The Alan Turing Institute

Request Allocation

Caller

Name	Martin O'Reilly
Telephone number	+44 0208 862 3359
Email	moreilly@turing.ac.uk
Department	Research Software Engineering

Project

Project title	QUiPP - Quantifying Utility and Privacy Preservation in synthetic data
Turing project code	R-SPET-202
Research programme	AI for science and government
PI/Supervisor name	Martin O'Reilly
PI/Supervisor email	moreilly@turing.ac.uk

Service

Which service	Azure
Subscription type	Project
New or existing allocation	New
Credits requested / $	95000
Start date	11 November 2019
End date	09 May 2021

Information

Data sensitivity	Sensitive
Platform justification	We are handling sensitive data sets and therefore need to use a Safe Haven. We will also have quite diverse and "peaky" compute requirements, so would also benefit from the flexibility provided by Azure. We anticipate requiring both large memory and GPU enabled machines at times.

The Alan Turing Institute

Research justification	We will be developing a benchmarking framework for synthetic data generation techniques to allow data holders to easily and safely evaluate for themselves how well various synthetic data generation techniques preserve the privacy and utility of their data, and how these trade off against each other. We will also provide benchmarking results across a range of sensitive datasets, and publish an analysis of how the performance of various techniques and their privacy-utility trade offs vary across different data types.
Computational requirements	We will need to run a Safe Haven for the 18 month duration of the project. We will need to run some computations on large memory and GPU enabled machines. Usage will be very "peaky" and we have estimated this as equivalent to 15 months of continuous compute on either a large memory or GPU VM. We will need to store a large amount of data. We're not sure how much data we will need to hold, but we anticipate several large datasets, including image data, so have costed 8 TiB of data. Note that this only represents an additional £360 / mo compared to a baseline 1 TiB storage capacity (£6,500 over the 18 months of the project), so is not our primary cost driver.
Users who require access and their emails	Safe Haven Production Administrators security group should have contributor or owner access and Safe haven Production Readers should have read access, as normal for Safe Haven deployments. No project members should have access to this subscription at first. This may change if and when devolved administration of compute VMs becomes available.

Costing

Costing breakdown	This is an ASG funded project.
	We are asking for funding at current costs, but have also included some indicative estimates of potential cost reductions should various planned Safe Haven improvements be successfully implemented. Any actual realised cost reduction will depend on whether and when such improvements become available and whether and when they can be retrofitted to an existing Safe Haven deployment.
	We are happy to be early adopters of these planned changes and help iron out any teething issues.
	At current costs ------------------ - Baseline Safe Haven with E16 compute VM for 18 months @ £2,000 / mo -> £36,000 - Budget to use either large memory or GPU VM for 15 mo @ £2,500 / mo -> £37,500
	Total cost: £73,500
	Potential reduced costs with Safe Haven VM consolidation plans --------------------- - Baseline Safe Haven with E16 compute VM for 18 months @ £1,600 / mo -> £28,800 - Budget to use either large memory or GPU VM for 15 mo @ £2,500 / mo -> £37,500
	Total cost: £66,300
	Potential reduced costs with Safe Haven VM consolidation plans + devolved VM administration rights permitting 12hr / day weekday only duty cycle --------------------- - Baseline Safe Haven with E16 compute VM for 18 months @ £1,000 / mo -> £18,000 - Budget to use either large memory or GPU VM for 15 mo @ £1,500 / mo -> £22,500
	Total cost: £40,500