Imperial College London

Summary of Key Research Data Environments

	Data Environment Name	Secure Enclaves	Big Data & Analytical Unit Secure Environment (BDAU SE)	Data Science Institute (DSI) Cloud Environment	ICT Research Computing Services
GENERAL DETAILS	**Description**	An environment for groups dealing with identifiable and de-identified sensitive data. It is designed to give stakeholders assurance that groups meet the Information Governance requirements that they are subject to (e.g. GDPR).	An ISO 27001 certified research environment, compliant with NHS Data Security and Protection Toolkit, located at the College data centre. It offers multi-user batch processing and an interactive GUI with database options, GPU/TPU facilities, and open source or proprietary software in a fully audited environment. The environment is suitable for ongoing research projects involving the use of de-identified personal data which needs to be handled in a secure way to ensure compliance with regulations such as GDPR, or the data security requirements as set by data providers.	Provides modern and future-proof cloud computing services for users and selected collaborators. Hosting of applications for processing sensitive data.	A multi-user batch processing environment, supporting a diverse set of high-throughput and high-end workloads. RCS provides bulk compute processing and high-volume data storage and processing capability to the College's research community.
	Owned By	Faculty of Medicine	Faculty of Medicine - BDAU	Data Science Institute	Central Faculty - ICT
	Hosted Location	On-premise (ICT Data Centre)	On-premise (ICT Data Centre)	On-premise (DSI private cloud environment)	On-premise (ICT Data Centre)
	Key Contact(s)	Mr Eric Johnson (eric.johnson@imperial.ac.uk)	BDAU Team: bdau@imperial.ac.uk	Dr Florian Guitton (f.guitton@imperial.ac.uk) Mr Oliver Nixon (o.nixon@imperial.ac.uk)	rcs-manager@imperial.ac.uk
KEY FACILITIES	**Facilities Offered**	• Each group is given a secured area of the College network (an enclave) that is isolated from all other areas of the network (including other groups in the environment). • Access to servers is through a controlled and audited management service. Users, workstations and servers are subject to a common Information Governance (IG) Policy. • Enclaves can contain servers running Linux or Windows operating systems. Servers may be physical, virtual or cloud servers, providing they meet the IG policy requirements. • BitLocker 256 encryption is used for data at rest on the server disks. Symantec desktop encryption is also used when sending data over the network. • The ICT backup service is used for all data stored by research groups.	• A standard operating/access model • Secure, fault tolerant data storage encrypted to AES256 • Open-source or proprietary analysis software • Multi-user batch processing • GPU / TPU facilities • Network speed: 10 Gbps • Ubuntu LTS based environment • Multi-layered multi-factor authentication	• Dedicated hardware for servers and network • Segregated physical network • Private virtual networks • Private virtual machines - user cloud environment • Fault-tolerant storage • Secure web applications • MongoDB database cluster • GPU cluster • TPU training and inference equipment • Open-source software and hardware solutions • Ubuntu LTS-based installation • Full hardware encryption • Network speed: 100 Gbps backbone, 20 Gbps minimum to the node.	• 80k compute cores in total, with the current maximum configurable job size of 2560 core with 40TB RAM • 160xRTX6000 GPU • 153 GPU of older specifications • Storage throughput of 25GB/s • Research Software Engineering support / consultancy / training available • Research Data Store service • Site-to-site data transfers with Globus web service
	Software Tools Provided (if any)	There is an R server within the Secure Enclaves. Research groups can install any required software applications on their own hardware connected to the Secure Enclaves.	• Jupyter • Python • R Studio • Stata • SPSS • SAS • MATLAB • Assorted user-coded software (e.g. Java, C, C++, C#)	DSI provides virtual environments where users have the freedom to install whichever piece of software they like.	Over 1,110 software packages registered, including: • MATLAB • SPSS • Gromacs • Openfoam • Star-ccm • Lammps • Gaussian • Comsol • Pytorch • R
	Environment Accessibility	**On-campus access:** internal users and external collaborators **Remote Access:** as above via VPN / Pulse Secure NB: Access to identifiable information is restricted to dedicated workstations. Access to de-identified information is possible through a dedicated College VPN connection.	**On-campus access:** internal users and external collaborators **Remote Access:** as above via VPN / Pulse Secure	**On-campus access:** internal users and external collaborators **Remote Access:** not available to users	**On-campus access:** internal users only **Remote Access:** not available via VPN or Pulse Secure
	Default Storage Allotment	No defaults. (Groups provide their own storage as part of the enclave servers. Data is backed up to ICT services.)	Varies by project (Typically starts at 4TB)	No defaults. (Storage is allotted on a case-by-case basis.)	1TB per user 2TB per project
	Maximum File Size	Not Applicable	The practical limit is the available storage at the time of file creation.	Not Applicable (Quotas are on a case-by-base basis)	Not Applicable
	Usable for Permanent Storage?	NO	NO	YES	YES (chargeable service)
	Data Backups?	NO	YES	This is possible, but not by default. Data backups require DSI action and access to a privileged account.	YES (chargeable service)
SECURITY	**Security-certified?**	YES	YES	NO	NO
	Certifications Available	ISO 27001 – Information Security Management NHS Data & Security Protection Toolkit	ISO 27001 – Information Security Management NHS Data & Security Protection Toolkit	NONE	NONE
	Suitable for Sensitive Data?	YES	YES (De-identified datasets only)	YES (De-identified medical data only)	NO
	Data Categorisations for Acceptable Datasets	Two types of enclaves are available: one for identifiable data; one for de-identified data. It is typical for groups to use both types. Servers in the identifiable enclave are subject to greater security controls.	Confidential Restricted Unrestricted	There are no limitations on data categories in this environment. Projects are responsible for understanding the legal implications. DSI does not enter into contractual agreements involving Imperial College as a legal entity.	Unrestricted
	Role-based Access Control?	YES	YES	YES	YES
COSTS	**Costs to Users**	Groups are required to purchase servers for processing in the enclaves. While no payment for enclave use is currently required, this will probably change.	**Environment Access:** £1,800 per year (free for IGHI staff) **Data Storage:** Free for projects with less than 1TB of data **Other:** Data transfer costs £300 (currently waived for a one-time transfer of data to the BDAU SE)	**Environment Access:** Free to users **Data Storage:** Free to users **Other:** Projects may make in-kind contributions to obtain dedicated resources.	**Environment Access:** Free at point of use **Data Storage:** Free up to 1 TB per user and 2TB per project. Researchers can use grant codes for additional resources.
REQUIREMENTS & RESTRICTIONS	**Access Requirements**	• Active College user account • Honorary contract (external researchers only) • Annual evidence of completed Imperial Essentials (data protection and information security awareness) training • Environment user registration • Evidence of completed DPIA / DART Registration • Dataset registration (with supporting documentation) • Users are responsible for acquiring all hardware	• Active College user account • Honorary contract (external researchers only) • Annual evidence of completed Imperial Essentials (data protection and information security awareness) training • Environment user registration • Dataset registration (with supporting documentation) • User and dataset registrations and training must be renewed in January every year	• Active College user account • Honorary contract (external researchers only) • Annual evidence of completed Imperial Essentials (data protection and information security awareness) training • Environment user registration • Evidence of completed DPIA / DART Registration • Dataset registration (with supporting documentation)	• Active College user account • Environment user registration
	Important Restrictions	• This environment is only for personal datasets. • Data can't be held within it for public access. • An Imperial College user account is needed. • Research supervisor advocacy required for student access • By default, users can't transfer data out of the environment. Approval is required for certain data files. • Groups will be required to provide evidence of key Information Governance requirements for ISO 27001 accreditation.	• This environment can't be used to store any identifiable data. • Data types not accepted: video or imaging data; audio data. • Data retention policy: research project duration only (subject to contractual data sharing agreements) • Research supervisor advocacy required for student access • Only aggregated research outputs (which are in line with the data provider's requirements) can be transferred out of the environment. • In exceptional cases, record-level data may be transferred out of the environment under an exception subject to the approval of the Director of the BDAU. • Data quality assurance, data curation, and general management for the data will be the responsibility of the Chief Investigator and their research team. • Users should not install their own software including any libraries/modules for existing software without prior authorisation of the BDAU.	• Research supervisor advocacy required for student access. • Projects are responsible for the security of their virtual environments. • DSI is not responsible for the upkeeping and updating of virtual environments used by projects. • Although they are possible, there are no default integrations with other College services.	• Data types not accepted: sensitive data • Research supervisor advocacy required for student access
FURTHER INFORMATION	**Information for Reference**	For more information on Faculty IG requirements: https://imperiallondon.sharepoint.com/sites/fom/operations/fomig/SitePages/FoM-Information-Governance.aspx	For more information about the BDAU and the Institute of Global Health Innovation (IGHI): https://www.imperial.ac.uk/centre-for-health-policy/our-work/data-science-and-analytics/big-data-and-analytical-unit-bdau/ http://www.imperial.ac.uk/global-health-innovation/	For more information about the Data Science Institute: http://www.imperial.ac.uk/data-science/	For more information about the ICT Research Computing Service or support requests: https://www.imperial.ac.uk/admin-services/ict/self-service/research-support/rcs/ https://www.imperial.ac.uk/admin-services/ict/self-service/research-support/rcs/meet-the-team/ https://servicemgt.imperial.ac.uk/ask

Imperial College – Data Classification Categories

C **CONFIDENTIAL DATA**
Sensitive personal data: the Special Categories of Data in Article 8 of the General Data Protection Regulation (GDPR). This includes patient identifiable data for research purposes. Additional protection measures are required.
Highly sensitive organisational data: data which could cause the College damage or financial loss if exposed, data protected by confidentiality agreements, legally privileged information, etc.

R **RESTRICTED DATA**
Personal data: any information related to a natural person which can be used to identify them (as defined in Article 4 of the GDPR). Special measures of protection are required.
Sensitive organisational data: this includes commercially sensitive planning / administrative / research data etc. Protection measures are required.

U **UNRESTRICTED DATA**
Non-personal data: organisational data.
Non-sensitive organisational data: data pertaining to the College which may or may not be published by default, but may be disclosed via Freedom of Information requests subject to legal advice.

Last Updated: 10/12/2021