Thames Valley & Surrey SECURE DATA ENVIRONMENT | Part of the NHS Research Secure Data Environment Network

Thames Valley and Surrey Secure Data Environment – Data Access Request Form

Ensure that all questions are answered. Guidance is provided for each question to assist you.

Projects approved for access to data within the Secure Data Environment (SDE) will be listed in our publicly available Data Use Register. Information for this listing will be extracted from the fields marked with the following icon: Do not include any confidential or sensitive information in these fields.

See UK Policy Framework for Health and Social Care Research - Health Research Authority for general guidance on the UK R&D landscape. If the proposed work is defined as research and you have completed an IRAS form, the relevant corresponding questions are indicated in the guidance text, where applicable. An index of all IRAS form questions, with guidance, is available here: IRAS Help - Reference - Collated Guidance - IRAS Form.

Details of person completing form			
Name		Role	
Organisation		Email address	

Safe Projects

1	Project Title	
		The title should: - identify the main area of your proposed work so that a member of the public would understand what you plan to study. - be the same as the title used on any other documents submitted for regulatory purposes (eg IRAS if submitted in the UK). **Do not include any confidential or sensitive information** as the Project Title will be publicly available on the Data Use Register if the application is successful.
		IRAS question A1
2	Lead Organisation	

		Provide the full name of the organisation leading the work and requesting access to data. This is the organisation signing the Data Access Agreement and taking responsibility for the named users on this specific request. **Do not include any confidential or sensitive information** as the Lead Organisation will be publicly available on the Data Use Register if the application is successful.	
3	Department within Lead Organisation		
		If your organisation does not have departments, leave blank.	
4	Other Organisation(s)		
		Any other organisation(s) involved with the work, including and/or employing individuals who will be accessing the data.	
5	a) Is the proposed work part of an existing study/project? b) If yes, who is the sponsor of that study (if not the Lead Organisation listed above)?	a) ☐ Yes ☐ No	b) Sponsor:
		Sponsor guidance: - UK Policy for Health and Social Care Research definition: The sponsor is the individual, organisation or partnership that takes on overall responsibility for proportionate, effective arrangements being in place to set up, run and report a research project. - The sponsor organisation must be validated by the SDE network before the proposed work can commence. - It is recognised that many projects will not have a separate sponsor to the Lead, ie contracting, Organisation.	
		IRAS question A64-1	
6	a) How is the proposed work defined? Use the HRA decision tool to establish this definition. b) If Research	a) ☐ Research ☐ Service evaluation/improvement/development ☐ Clinical/non-financial audit ☐ Health surveillance ☐ Other – specify:	b) i) ☐ Yes ☐ No, but I intend to apply ☐ Not required – justification: ii) IRAS ID:
	i) Does the proposed work have HRA approval, where required?	a) See the Is my study research? tool, and defining research table, for further guidance. b) HRA approval is not the same as REC approval. See HRA approval: HRA Approval - Health Research Authority and HRA Data Decision Tool Start - HRA Data Decision Assessment Tool for further guidance.	

	If HRA approval is not required, explain why ii) IRAS ID	
7	Which use cases will be addressed by the proposed work? Choose all that apply. If AI/algorithm development, what stage is it at? Choose all that apply.	Real world studies ☐ Clinical effectiveness and treatment outcomes ☐ Compliance with current prescribing limits ☐ Cost effectiveness ☐ Outcomes based agreements ☐ Pharmacogenetics ☐ Pharmacovigilance and safety ☐ PROMs based research ☐ Regulatory Submissions ☐ Verification and validation of product performance
		Epidemiological studies ☐ Burden of disease and unmet need ☐ Genomic epidemiological research ☐ Natural history of disease ☐ Risk factor identification ☐ Social determinants of health and health inequalities
		Health systems research ☐ Adherence to guidelines ☐ Health service research and health care quality ☐ Learning health systems ☐ Pathway mapping
		Clinical trial activities ☐ Feasibility of clinical trials ☐ Follow-up and long-term outcomes ☐ Patient recruitment and enrolment ☐ Virtual controls and single arm trials
		Translational research

		☐ Biomarker development ☐ Biomarker research ☐ Creating integrated risk scores ☐ Identification of drug targets and drug discovery ☐ Precision medicine/genomic profiling	
		AI/algorithm development ☐ Natural language processing in healthcare ☐ Predictive analytics for healthcare management ☐ Prognostic predictors of outcomes ☐ Risk stratification and case finding ☐ Validation of AI algorithms for diagnosis and treatment	Stage ☐ Research ☐ Development ☐ Validation
		These use cases are how the proposed work is classified in the SDE Network.	
		IRAS question A7	
8	a) Is the proposed work intended to support the development of a particular drug or technology? b) If yes, what stage is it at?	a) ☐ Yes ☐ No	b) If yes: ☐ initial discovery and development ☐ regulatory approval ☐ post-market surveillance
9	What health condition, disease and/or therapeutic areas does the proposed work focus on?	☐ Blood ☐ Cancer and neoplasms ☐ Cardiovascular ☐ Congenital disorders ☐ Ear ☐ Eye ☐ Infection ☐ Inflammatory and immune system ☐ Injuries and accidents ☐ Mental health ☐ Metabolic and Endocrine	

		☐ Musculoskeletal ☐ Neurological ☐ Oral and Gastrointestinal ☐ Renal and Urogenital ☐ Reproductive health and childbirth ☐ Respiratory ☐ Skin ☐ Stroke ☐ Generic health relevance – specify: ☐ Disputed Aetiology and Other – specify:	
		Use the Health Research Classification System (HRCS) Category (primary). The Health Categories dimension of the HRCS captures the area of health or disease being studied. There are 21 separate categories which encompass all diseases, conditions and areas of health. Each of the Health Categories includes research into both disease and normal function. Of the 21 categories, 19 refer to specific areas of health or disease. The Generic Health Relevance category has been included to capture research that is relevant to all diseases and conditions or to general health and well-being. The Disputed Aetiology and Other category is included to code research that does not fit within the Generic Health Relevance category or the 19 health specific categories.	
10	a) What is the funding status of the proposed work? b) Give details of funding arrangements	a) ☐ Funding secured ☐ Funding applied for ☐ Funding application being prepared ☐ Need help with funding application	b) Funding details:
		The SDE team may be able to support with funding applications, and we welcome enquiries and discussions regardless of funding application stage. Include details of the stage of and deadlines for any grant applications, or anything else you think will be helpful for the assessment of your request. Funding must be in place before access to data can be granted.	
		IRAS question A65	

11	a) Source of funding applied for/secured. b) Name of organisation(s) providing/being targeted for the funding.	a) ☐ Academic Research Institutions ☐ Government Agencies, eg DHSC, Regulatory bodies, NICE ☐ Healthcare Providers ☐ Non-Profit Organisations, eg foundations, advocacy groups, charities ☐ Other for-profit organisations (SME) ☐ Other for-profit organisations (non-SME) ☐ Pharmaceutical and Biotechnology Companies (non-SME)	b) Name of funder(s):
12	Type of data access required - select one option only.	☐ Access to deidentified data in a project workspace: record level data ☐ No record level data access required – only algorithm/query: aggregated data only ☐ Additional data for existing, consented cohort: record level data, contains individual identifiers	
		NB: The standard type of access offered by the SDE is access to deidentified data in a project workspace: record level data.	
13	Proposed start date for data access (dd/mm/yy).		
		The proposed start date for accessing the requested data, rather than the start date of any broader, related programme of work. This date cannot be in the past.	
14	How long do you expect to require data to be available for analysis within the SDE?	☐ up to six months ☐ up to one year ☐ up to two years ☐ up to three years ☐ up to four years ☐ up to five years	
		It may be possible to extend the time period should activity take longer than anticipated. If you expect to require access for more than five years, you will need to renew your application before five years comes to an end, in line with HRA guidance. **This information will be added to the Data Use Register** and will be agreed as part of your Data Access Agreement.	
		IRAS question A69-1	

<table>
<tr><td rowspan="2">15</td><td rowspan="2">a) Has the proposed work been submitted for ethical review by an independent committee and/or expert(s)?

b) If yes, give details, ie by whom, type of review, outcome and any other information you think is useful; provide evidence, ie review outcome report, REC number if obtained etc.

If no, the SDE will review the proposed work under its REC research database approval.</td><td>a) ☐ Yes ☐ No</td><td>b) If yes, please attach IRAS form and approval letters to this application:

By whom:

Type of review:

Outcome:

Other information (e.g. REC Reference):

If no:

☐ I confirm I understand that the SDE will make an assessment of each request in accordance with REC Research Database Conditions.</td></tr>
<tr><td colspan="2">a) This question is intended to provide context of the broader assessment and review processes the proposed work has already been through and will seek to reduce duplication for users. Assurance is being sought that the thinking regarding the proposed work is mature and it has already been assessed on areas like methodology. It is recognised that not all sectors practice this kind of review regularly.

b) The SDE access review process is designed in line with the REC Research Database Condition. If there has been no ethical review the SDE team will make an assessment of each request in accordance with REC Research Database Condition Option A: Generic approval for external researchers.

The SDE team will review the request from an ethical perspective, looking specifically at the questions in this form that cover:

- aims and use case of proposed work
- potential benefit
- patient and public involvement
- data requirements and rationale
- risks and mitigations
- potential bias, including selection and casual
- how results and outcomes will be shared
- the proposed legal basis
- conflicts of interest</td></tr>
</table>

16	a) What is your proposed legal basis for requesting that the SDE processes personal data and provides access to a dataset for the proposed work? If you are not using the legal basis used by the SDE, select Other. b) If you have selected Other above, specify the legal basis from: i) Common law duty of confidentiality ii) GDPR Article 6 iii) GDPR Article 9	a) ☐ The legal basis used by the TVS SDE (see guidance below for TVS SDE legal basis) ☐ Other, provide details and attach approvals to application:
		b) If Other (pick one from each) provide details and attach approvals to application: : i) Common law duty of confidentiality ☐ Consent ☐ Section 251 exemption ☐ N/A ii) GDPR Article 6 ☐ 1(a) Consent ☐ 1(b) Necessary for the performance of a contract to which the data subject is or about to be party ☐ 1(c) Necessary for compliance with legal obligation ☐ 1(d) Necessary to protect the vital interests of the data subject ☐ 1(e) Necessary for performance of a task carried out in public interest or in exercise of official authority ☐ 1(f) Legitimate interest (does not apply for public authorities)up to three years iii) GDPR Article 9 ☐ 2(a) Explicit consent ☐ 2(b) Necessary in connection with employment ☐ 2(c) Necessary to protect the vital interests of the data subject ☐ 2(d) Legitimate interest ☐ 2(e) The data subject has manifestly made the information public

		☐ 2(f) Necessary for establishment, exercise or defence of legal claims ☐ 2(g) Necessary for reasons of substantial public interest ☐ 2(h) Necessary for provision of health and/or social care, including preventative or occupational medicine ☐ 2(i) Necessary for reasons of public interest in the area of public health ☐ 2(j) Necessary for archiving purposes in the public interest, scientific or historical research or statistical purposes.
		You are required to understand the lawful basis under which you are making this request. The common law duty of confidentiality is satisfied through a Section 251 approval from the Health Research Authority following advice from their Confidentiality Advisory Group. The operation of the SDE meets the requirements of the UK GDPR and Data Protection Act 2018; in particular, the following conditions from Articles 6 and 9: - Article 6 (1) (e): processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority. - Article 9 (2) (j): processing is necessary for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes. Schedule 1, Part 1 of the Act states that Condition 9 (2) (j) is met if the processing: - is necessary for archiving purposes, scientific or historical research purposes or statistical purposes, - is carried out in accordance with Article 89(1) of the GDPR and is in the public interest.
17	Any additional information that has a bearing on the assessment of the proposal that should to be treated in confidence.	
		Use this field to share any information that you consider to be commercially or otherwise sensitive that you feel is important for the assessment of the request. Any information provided here will be treated in confidence, in line with SDE confidentiality agreements, ie not published on website or shared beyond the SDE team, and SARC where appropriate.
18	Benefits – part 1 Who stands to potentially benefit from the use of this data?	☐ Individual patients, carers or families – details: ☐ The population in general or specific communities – details: ☐ The health and care system – details: ☐ Health and care staff – details: ☐ Society, including economic benefits and/or benefits to industry/companies – details: ☐ Other – details:

	Choose at least one, and all that apply. Describe in what ways for each of the chosen options - read the guidance carefully to inform your answer.	Describe the potential benefit you anticipate for each option. Include any measurable outcomes, examples, or evidence if available. **Individual patients, carers or families** - does the use of data address what matters to patients and their families? Could the use of this data: • improve care and/or the experience of it? • improve or create new treatments? • Improve outcomes eg health, quality of life, happiness? **The health and wellbeing of the general population and communities** Could the use of this data benefit the whole population and/or specific communities? Could it: • improve knowledge about disease/illness eg aetiology, prevalence, incidence, morbidity, mortality etc? • improve health equity/reduce health inequalities? • improve healthy living or quality of life? • reduce risk factors eg in the environment? **The health and care system** Could the use of this data improve: • quality – effectiveness, safety and/or experience? • efficiency, including productivity or investment? • pathway design and/or delivery? • standards of care? • demand or pressure on local health services? **Health and care staff** Could the use of this data improve staff: • experience? • knowledge? • skills? • wellbeing? • recruitment and/or retention? **Society**

		Could this use of data generate non-health benefits for society for example: • Economic value, eg increase in jobs • investment with money flowing back into the local system/area? • value generation for companies eg investment or revenue increases? • Trust and confidence in the NHS or in public services more broadly? • Research value eg new jobs, investment, trust? • Non-health benefits, eg on crime or education?
19	Benefits – part 2 More specifically, what kind of benefit could the use of this data potentially deliver? Choose at least one, and all that apply.	☐ Further understanding of the health and care needs of populations ☐ Lead to the identification and progress of treatments and therapies to treat illness ☐ Further understanding of regional and national trends in health and social care needs ☐ Address healthcare inequalities ☐ Support the quality and safety of services ☐ Inform planning health services and programmes ☐ Inform design of prevention interventions and evaluation interventions ☐ Inform decisions on how to effectively allocate and evaluate funding according to health needs ☐ Other – specify:
		The concept of public benefit is broad and flexible, and could include direct, indirect, and long-term benefits. The benefit needs to be identifiable regardless of whether it can be quantified or measured.
		IRAS question A24
20	a) Describe what patient and public involvement has already been undertaken. b) Describe what patient and public involvement is planned.	a) PPIE undertaken: ☐ Yes ☐ No If Yes, complete all of the below: What sorts of people were involved, eg patients with myeloma, people from Asian communities: How many people have been involved: What you did with them, eg workshop/focus group: What difference did involvement make to this proposal: If No: Details of why no PPIE undertaken:

		b) PPIE planned: ☐ Yes ☐ No If Yes, How will you involve people in defining the need/questions to be answered; governance; design; delivery; dissemination; evaluation: What sorts of people do you plan to involve: How many people do you plan to involve: What difference do you expect this involvement to make: If No: Details of why no PPIE planned:
		While patient and public involvement is an expectation for the majority of requests, it is acknowledged that there are situations where these activities are not appropriate. If no public involvement has taken place or is planned, explain why. Requests that are able to adequately justify why patient and public involvement is not appropriate can be progressed. Documentation previously prepared for grant applications (ie NIHR) may be submitted as part of the answer to this question. The SDE team can support you in identifying whether patient and public involvement is required and potential further PPIE options – such support may have additional cost implications. Further guidance: - Principles of PPIE: Public Involvement – Health Research Authority - PPIE planning tool: https://plan4ppie.com. - How best to present PPIE evidence in an application: IRAS Help: Preparing & submitting applications – Public Involvement
		IRAS questions A14-1, A6-2, A13, A22, A30-1, and A51.
21		**Risk 1** Description: Mitigation: **Risk 2** Description: Mitigation:

		Risk 3 Description: Mitigation: **For additional risks, use this box to add as many as required, ensuring you include all the fields above for each:**	
		A data use purpose cannot be considered as beneficial if the harm resulting from the use of the data outweighs the public benefit. Describe how you have taken into consideration the potential risks and benefits and mitigations in the design of the proposed work. Consider: - Could individual patient privacy be compromised? For example, consider whether the user holds any separate data which might enable reidentification of individuals, even if this data will not be brought into the SDE. - Could patient or service user safety be harmed? - If data was to be used for this purpose, could it make some patients or service users less likely to seek care or be less frank in discussion with health and care professionals? - Could the purpose lead to the creation or exacerbation of inequalities or unlawful discrimination against particular communities? - Could the use of inaccurate or inadequate health and social care data lead to unrepresentative findings? - Could the purpose undermine the sustainability of publicly funded health or adult social care services?	
		IRAS questions A22 and A23	
22	a) Do you have access to any data that might increase the risk of you being able to reidentify any individuals within the data provided for this request? b) If yes, identify the data you have access and how you will	a) ☐ Yes ☐ No	b) If yes:
		For example, the user may have access to identifiable data as an employee of a data provider.	
		IRAS questions A38, A39 and A40	

	mitigate the risk of reidentification.	
23 	Summary of the proposed work, written in **Plain English**. This should include brief descriptions of: a) Aims: what the proposed work is trying to achieve/show/prove/ develop  b) Methodology and proposed analysis c) What data is requested d) Inclusion and/or exclusion criteria e) Duration of proposed work f) Potential value of the proposed work – benefits to patients, NHS, society etc (see guidance for Q18)  g) Any other information	a) Aims:
		b) Methodology and analysis:
		c) Data requested:
		d) Inclusion and exclusion criteria:
		e) Duration:
		f) Potential value:
		g) Any other information:
		The information requested for this summary is also requested in more detail in other questions. The aim of this question is to create a **Plain English** summary that is suitable for a non-expert, public audience, and that can be published. **Do not include any confidential or sensitive information** as the Aims and Potential Value will be publicly available on the Data Use Register if the application is successful. The summary should:

		- use the headings provided - be in language suitable for non-experts, and for members of the general public - explain all technical or complex terms - not use jargon, acronyms or abbreviations - aim for a reading age of 9 – 11 - use short sentences The SDE team will review and request iterations if additional information/clarification is required. For more guidance see: - Plain English Medical Information Guide - IRAS Guidance – What to include in the research summary - NIHR Plain English summary guidance
		IRAS questions A7, A12, A13, A17-1, A17-2, A69-1

Safe Data

24	a) Which patients do you require data for, ie what must be true of a patient for them to be included in/excluded from this cohort? b) Are the any sub-cohort requirements? If yes, what are they? c) How have you identified this cohort/sub-cohorts and what logic/rationale have you used to ensure that it will support the proposed work whilst	a)	b) ☐ Yes ☐ No If yes:
		c)	
		a) and b) A description of the criteria that define the patients to be included and to be excluded from the requested data. These might include age, demographics, volume, date ranges, diagnostic codes (this should include ICD10/SNOMED codes if known), or procedure codes. Explain which date field(s) in the data specification will be used to define the requested cohort (eg date of admission or date of operation). This question is likely to require iteration and discussion with the SDE team.	

	also being proportionate?	c) This question aims to understand what methodology was used to identify the required cohort. Significant detail on statistical analysis plans is not required, rather a high-level rationale is sought for why this cohort has been requested. This question also seeks to establish how data minimisation has been considered, ie that the proposed criteria provide sufficient data to enable the proposed work to be conducted without including unnecessary information/data.	
		IRAS question A17-1, A17-2 and A60	
25	Approximately how many patients are required for the proposed work?		
		Anticipated/required cohort size, based on your understanding of the proposed work – this will allow the SDE team to assess if there is data on a sufficient number of patients within the SDE.	
		IRAS question A59	
26	a) What kinds of data do you need for the proposed work? b) Give further details of your specific requirements to support your selections in a). c) Explain briefly why these kinds of data are needed.	a) Coded and numeric: ☐ Nationally-reported data, including commissioning data sets ☐ Clinical data in hospital systems Free text and documents: ☐ Pathology reports ☐ Radiology reports ☐ MDT (multi-disciplinary team) meeting reports ☐ Other reports, letters, or documents– specify:	b) Further details:

		Medical imaging: ☐ X-ray ☐ CT or MRI ☐ Ultrasound ☐ Other imaging– specify: Other care settings: ☐ Mental health ☐ Primary care ☐ Ambulance services ☐ Other – specify:	c) Justification:
		a) Detailed data specifications will be worked out with the SDE team following this application. Data held in free text or document form will be abstracted or redacted as appropriate. If kinds of data you require are not listed, choose 'Other' and specify. List here if you require data from a site not within the SDE region. The SDE team may be able to support such requests. b) This question aims to understand how you have identified what data is required. Significant detail on statistical analysis plans is not required, rather a high-level justification and rationale for why this data has been requested. This question also seeks to establish how data minimisation has been considered, ie that the proposed criteria provide sufficient data to enable the proposed work to be conducted without including unnecessary information/data.	
		IRAS question A60	
27	Given the patient cohort and data requested, how will be any potential sources of bias, eg ethnicity data, health equity/inequalities be accounted for and managed?		
		Selection bias is a particular problem inherent in case-control studies, where it gives rise to non-comparability between cases and controls. In case-control studies, controls should be drawn from the same population as the cases, so they are representative of the population that produced the cases. Also consider causal and other types of bias. Further guidance: - Biases and Confounding | Health Knowledge - FOR-EQUITY – tools and resources to help reduce social and health inequalities	

28	Date range of requested data extract: a) Start date (yyyy-mm-dd) b) End date (yyyy-mm-dd)	a) Start date:	b) End date:
		a) If day and month are not relevant, choose the first day and month of the year. b) If day and month are not relevant, choose the last day and month of the year.	
29	a) Will any refreshes of the data be required? b) If yes, what kind (periodic or scheduled)? i) If periodic, state frequency ii) If scheduled, state kind iii) Justification for refresh requirement and kind/frequency	a) ☐ Yes ☐ No	b) If yes: i) If periodic: ☐ Monthly ☐ Quarterly ☐ Bi-annually ☐ Annually ii) If scheduled: ☐ Milestone-driven ☐ Define milestone: ☐ Other – specify: iii) Justification:
		Examples: - periodic, monthly: every 30 days to meet the monthly review cycles described in the project - periodic, bi-annually: every 6 months to account for seasonal changes in hospital admissions - scheduled, milestone-driven: for final analysis	
30	a) Are any additional software, applications, tools **beyond** the standard offer required in the analysis environment that are	a) ☐ Yes ☐ No	b) If yes: Software/application/tool name(s): Link(s)/description(s): Justification:

	essential to conduct the proposed work? b) If yes: - identify the software/ application(s)/ tool(s) - provide a link to identify the software/ application(s)/tool(s) where available, otherwise provide a detailed description of the software/ application(s)/tool(s) - provide a justification	The SDE provides a standard software, application, tool offer, available to all users with approved requests: - Python - Anaconda - Visual Studio Code - R - RStudio - Azure Storage Explorer - Git - LibreOffice - Package Mirror Provide a clear description and justification for any additional requirements to enable the SDE to assess your request, noting that not all requirements can be accommodated. Providing a link/detailed description will enable the SDE team to make a faster and more accurate assessment of your request. If you are not able to define whether additional software/application(s)/tool(s) are needed at this stage, it may be possible to add further software at a later stage of the proposed work. Licensing for additional software/application(s)/tool(s) may have cost implications for the proposed work.

31	Compute requirements for the proposed work. Choose all that apply and indicate estimated usage in hours per month for the chosen options.	**Series**	**vCPU**	**RAM (GB)**	**GPU**	**GPU Memory (GB)**	**CPU Type**	**Estimated usage – hours per month**
		CPU only VMs:						
		☐ B Series	2	4	None	-	Intel	
		☐ Dsv5 Series	2	8	None	-	Intel	
		☐ D4as_v5 Series	4	8	None	-	AMD	
		☐ **Fs_v2 Series***	4	8	None	-	Intel	
		☐ Dsv5 Series	4	16	None	-	Intel	
		☐ Fs_v2 Series	8	16	None	-	Intel	
		☐ Dsv5 Series	8	32	None	-	Intel	
		☐ Easv4 Series	8	64	None	-	AMD	
		☐ Fs_v2 Series	16	32	None	-	Intel	
		☐ Dsv5 Series	16	64	None	-	Intel	
		GPU VMs - A10:						
		☐ A10_v5 Series	6	55	1/6 A10	4	AMD	
		☐ A10_v5 Series	12	110	1/3 A10	8	AMD	
		☐ A10_v5 Series	18	220	1/2 A10	12	AMD	
		☐ A10_v5 Series	36	440	1 A10	24	AMD	
		☐ A10_v5 Series	72	880	2 A10	48	AMD	
		GPU VMs – H100:						
		☐ H100_v5 Series	40	320	1 H100 NVL	94	AMD	
		☐ H100_v5 Series	80	640	2 H100 NVL	188	AMD	
		The SDE provides a standard compute offering as above (see Appendix one for example use cases for each option), available to all users with approved requests. *** If you're unsure of your requirements at this stage, select the DEFAULT offering**: - Fs_v2 Series | 4 CPU | 8GB RAM | Use case: CPU-intensive computations and model inference Estimated usage time will be used to calculate indicative costs – additional usage is permitted and will be charged according to the schedule agreed at contracting.						

<table>
<tr><td rowspan="2">32</td><td rowspan="2">a) Are there any compute requirements <u>beyond</u> the standard offer in the analysis environment that are essential to conduct the proposed work?

b) If yes:
- identify the compute requirement(s)
- provide a justification</td><td>a) ☐ Yes ☐ No</td><td>b) If yes:
Compute requirement(s):
Justification:</td></tr>
<tr><td colspan="2">Provide a clear description and justification for any additional requirements to enable the SDE to assess your request, noting that not all requirements can be accommodated.

If you are unsure whether there are additional compute requirements at this stage, it may be possible to add further requirements in the future.

Additional compute requirements may have cost implications for the proposed work.</td></tr>
<tr><td rowspan="2">33</td><td rowspan="2">a) Are there any storage requirements in virtual machines <u>beyond</u> the standard offering in the analysis environment that are essential to conduct the proposed work?

b) If yes:
- identify the storage requirement(s)
- provide a justification</td><td>a) ☐ Yes ☐ No</td><td>b) If yes:
Storage requirement(s):
Justification:</td></tr>
<tr><td colspan="2">The SDE provides a standard, DEFAULT storage offering, available to all users with approved requests:

- Shared workspace storage: 200Gb Azure Files
- Windows: 128Gb disk
- Linux: 30Gb disk

Provide a clear description and justification for any additional requirements to enable the SDE to assess your request, noting that not all requirements can be accommodated.

If you are unsure whether there are additional storage requirements at this stage, it may be possible to add further requirements at a later stage of the proposed work.

Additional storage requirements may have cost implications for the proposed work.</td></tr>
<tr><td>34</td><td>a) Do you require any code to be uploaded to</td><td>a) ☐ Yes ☐ No</td><td>b) If yes:
Code details:
Justification:</td></tr>
</table>

	the analysis environment? b) If yes - give details of the code - provide a justification	All code must comply with the Analysis Environment Terms of Use. The Lead Organisation (Q2) is responsible for ensuring that any requirements and permissions for using this code are met, eg licensing restrictions.
35	a) Do you require your own dataset, from outside the SDE, to be integrated with data within the SDE? b) If yes, provide details on: i) Type (eg experimental results, user data) and method of collection ii) Format iii) Volume/size, eg 10,000 records, 500 MB iv) Deidentification type v) Is linkage of your own data to data being provided by the SDE required? If Yes, provide justification, and discuss with requirements with	a) ☐ Yes ☐ No b) If yes: i) Type and collection method: ii) Format ☐ CSV ☐ TSV ☐ Excel ☐ JSON ☐ SQL database ☐ Other – specify: iii) Volume/size: iv) Deidentification type: ☐ anonymised/aggregate ☐ pseudonymised ☐ identifiable ☐ reference v) ☐ Yes ☐ No If yes: Justification: ☐ I confirm I have discussed data linkage requirements with the SDE team vi) Permissions:

	SDE team prior to submitting this request. vi) Describe the permission you have for using this data for the proposed work (eg written agreement from data owner, REC/CAG) c) What is the legal basis for processing this dataset for the proposed work: how does your data comply with relevant regulations and standards (ie GDPR, common law duty of confidentiality)?	c) Legal basis: GDPR Article 6 ☐ 1(a) Consent ☐ 1(b) Necessary for the performance of a contract to which the data subject is or about to be party ☐ 1(c) Necessary for compliance with legal obligation ☐ 1(d) Necessary to protect the vital interests of the data subject ☐ 1(e) Necessary for performance of a task carried out in public interest or in exercise of official authority ☐ 1(f) Legitimate interest (does not apply for public authorities)up to three years GDPR Article 9 ☐ 2(a) Explicit consent ☐ 2(b) Necessary in connection with employment ☐ 2(c) Necessary to protect the vital interests of the data subject ☐ 2(d) Legitimate interest ☐ 2(e) The data subject has manifestly made the information public ☐ 2(f) Necessary for establishment, exercise or defence of legal claims ☐ 2(g) Necessary for reasons of substantial public interest ☐ 2(h) Necessary for provision of health and/or social care, including preventative or occupational medicine ☐ 2(i) Necessary for reasons of public interest in the area of public health ☐ 2(j) Necessary for archiving purposes in the public interest, scientific or historical research or statistical purposes. Common law duty of confidentiality ☐ Consent ☐ Section 251 exemption ☐ N/A

		The Lead Organisation (Q2) is responsible for ensuring appropriate permissions and approvals are in place for all external data. b) Data description: v) Discuss requirements for linkage with the SDE team prior to submitting this request vi) Include permissions, compliance, lawful basis and minimisation to avoid risks. List data owner if applicable and upload agreement giving permission to use this data for this proposed work. If linking this data to other data within the SDE would render the data re-identifiable to the user (using any means reasonably likely to be available to them), the Lead Organisation (Q2) must demonstrate that: - the common law duty of confidentiality has been satisfied; and - lawful bases are in place from UK GDPR Articles 6 and 9. Additional REC approval will also be required, as well as Section 251 support via CAG if explicit consent for accessing identifiable data for research purposes has not been given. Re-identifiability may become possible because the data accessed are linked with other data about the same individual or because of unique outliers created permitting identity inference. UK GDPR compliance will be applicable. For further guidance, see Can we identify an individual indirectly from the information we have (together with other available information)?.	
36	a) Do you require the workspace to be archived at the end of the proposed work beyond the standard SDE archiving policy? b) If yes, how long do you expect to require the workspace to be archived post the end of the proposed work beyond the standard archiving policy?	a) ☐ Yes ☐ No	b) If yes: ☐ seven plus one year ☐ seven plus two years ☐ seven plus three years ☐ seven plus five years ☐ longer – specify: seven plus years
		Standard workspace archiving policy: seven years. This includes the VM disks, approved airlocks, and the shared storage space. Any additional time the workspace is required to be stored for or archived following the close of the proposed work to allow for clarifications, clinical inspection etc. before publication. If you require additional time, further clarification may be sought regarding any activities you propose to undertake in this time period. The workspace may be stored by the SDE for longer than the period required by the proposed work if there are other retention obligations, eg Retention of Trial Records.	

		IRAS questions A43, A44, A45
37	Outline the proposed analysis methodology (in Plain English), including relevant statistical and analytical techniques and key variables.	
		This question aims to understand the wider context of the proposed work and data use. The SDE undertakes scientific critique of the information science in the context of the actual data being requested. The SDE team and Services and data Access Review Committee (SARC) is made up of members who may not be experts in your specific area of work. When answering this question you should use language appropriate for non-experts (ie Plain English), so all members are able to understand the content. It is recognised that methodology may evolve over the duration of the proposed work. The SDE team and SARC will not use this information to judge the validity of the methodology and statistical and analytical techniques; the information is requested to allow the team and committee to get a better understanding of the proposed work as whole.
		IRAS question A13, A62
38	Given the proposed methodology, how will be any potential sources of bias, eg ethnicity data, health equity/inequalities be accounted for and managed?	
		While selection bias is a particular problem inherent in case-control studies, where it gives rise to non-comparability between cases and controls, it should be considered in all types of studies. Also consider causal and other types of bias. Further guidance: - Biases and Confounding | Health Knowledge - FOR-EQUITY – tools and resources to help reduce social and health inequalities
Safe Outputs		
39	What outputs do you expect to export from the SDE? For each output: - describe the output - describe how it will be used during and	**Output 1** Output description: Use description: Export frequency: ☐ Weekly ☐ Monthly

	after the project has closed - define the frequency of export required - explain why you require the chosen frequency	☐ Quarterly ☐ End of project only Frequency description: **Output 2** Output description: Use description: Export frequency: ☐ Weekly ☐ Monthly ☐ Quarterly ☐ End of project only Frequency description: **Output 3** Output description: Use description: Export frequency: ☐ Weekly ☐ Monthly ☐ Quarterly ☐ End of project only Frequency description: **For additional outputs, use this box to add as many as required, ensuring you include all the fields above for each:**
		Outputs must not be row-level data; only aggregate results of analysis and code, subject to approval, will be permitted to be exported.

		It is acknowledged that this may evolve during the proposed work - at this stage an indication of the types of proposed outputs is required to enable assessment of the request. If the request is approved, further checks will take place on actual outputs, and they may be subject to further approvals/agreements. Approval of a request using this form does not constitute approval of actual project output export requests. The Data Access Process will consider what the outputs of the proposed work will look like and make sure that: - there is no risk to the privacy of patients. - data can only be removed with permission. - If required, there are additional conditions set on onwards publication and use.
40	How will the results and outcomes of the proposed work be published and/or disseminated to stakeholders and interested parties, including patients and the public?	

		The public and decision makers may not read the scientific/academic literature, or attend conferences where research is presented. Describe how the results and impact of your proposed work will be shared broadly, including describing the public benefits of the proposed work. This could include press releases, social media, progress reports on websites that the public might access and/or feedback to patient groups.
		
		It is good practice to ensure that those patients and the public involved in your work (either as participants or as part of your PPIE activities) are informed about the results and impact.
		
		If your request is approved, a yearly written statement to describe project progress is required. Once the proposed work has been completed, or has otherwise ended, an impact statement will be requested. Both the annual updates and impact statement will be published on the SDE website as part of the Data Use Register.
		
		Information for impact statement:
		
		- Project name
		- Organisation and individuals involved (with contact details)
		- Data accessed
		- Intended outcomes
		- What happened in practice
		- Describe impact on:
		  ○ patients, carers, families
		  ○ general population and/or specific communities
		  ○ the health and care system
		  ○ health and care staff

		○ society more broadly – non-health benefits ○ economic indicators - Dissemination plan: ○ publications ○ conferences ○ further research ○ policy development The HRA has issued guidance on making health and care research findings accessible: - Research Transparency – setting out expectations for sponsors, researchers, and funders - Writing a plain language (lay) summary of your research findings
		IRAS questions A51, A52
Safe People		
41	Contact person for this request (proposal, contracting, changes to the project details and data specification)	
		You can select from the list of contacts listed in the organisation's registration (this can be updated by the organisation's authorised users if necessary) – you can add more than one contact; if you do, indicate who is the primary contact.
42	Contact person to approve in-project requests (budget, additional users) during project delivery	
		You can select from the list of approvers listed in the organisation's registration (this can be updated by the organisation's authorised users if necessary) – you can add more than one approver; if you do, indicate who is the primary approver for these purposes.
43	List all users who will require access to the project workspace. If all team members have been identified and appointed at this stage, list all individuals by full name job title	
		You can select from the list of people listed in the organisation's registration (which the organisation main users/contacts can update if necessary) – you can add more than one person. You should only list individuals who will be interacting with the data, recognising that on large or long-term projects there may be many individuals involved in the programme who do not come into contact with the data. Only users who are validated with the SDE network will be able to access the data.

	and employing organisation. If individuals have not yet been appointed, describe the anticipated project team that will access and analyse the data by job title, role description and employing organisation.	If a project team has not yet been appointed, you will be asked to confirm names and email addresses at a later stage of the application process. Additional users and team amendments can be made later in the process.
44	a) In relation to the proposed or subject matter arising from the specific individuals named in this request, are there any known conflicts of interest? b) If yes, for each conflict of interest, define/describe: - the category - the conflict of interest - associated risks - associated mitigations	a) ☐ Yes ☐ No b) If yes, for each: **Conflict of interest 1** Category: ☐ Direct financial ☐ Direct non-financial ☐ Indirect Description: Risks: Mitigations: **Conflict of interest 2** Category: ☐ Direct financial ☐ Direct non-financial ☐ Indirect Description:

		Risks: Mitigations: **Conflict of interest 3** Category: ☐ Direct financial ☐ Direct non-financial ☐ Indirect Description: Risks: Mitigations: **For additional COIs, use this box to add as many as required, ensuring you include all the fields above for each:**
		Identify any personal conflicts or interests that the SDE team should be aware of in the context of this request, and associated mitigation. Interests are categorised as: - Direct financial (eg employment, consultancy, shareholdings) - Direct non-financial (eg advocacy roles, professional memberships) - Indirect (eg interests of close associates) Some aspects to consider (not exhaustive): - Could individual patient privacy be compromised? This could happen if a user has access to any other data that might enable reidentification of individuals, even if this data will not be brought into the SDE, eg do any users have access to clinical systems as part of their employment (ie if anyone accessing the requested data for the proposed is also an employee of any of the provider organisations). - Does the Host Organisation (for TVS SDE this is OUH NHS FT), or any of the other provider organisations, have an equity stake in the Lead Organisation (Q2)?
45	a) Are any individuals or organisations who will access the data	a) ☐ Yes ☐ No

	located outside the UK or EU, or subject to non-EU data protection laws? b) If yes, who, and where will they be located when accessing the data?	b)

Appendix one

Standard Compute Offering

Series	vCPU	RAM (GB)	GPU	GPU Memory (GB)	CPU Type	Use Case
CPU only VMs						
B Series	2	4	None	-	Intel	Basic coding, text editing, simple scripts
Dsv5 Series	2	8	None	-	Intel	Data preprocessing, small dataset analysis
D4as_v5 Series	4	8	None	-	AMD	Statistical analysis, medium data processing
Fs_v2 Series	4	8	None	-	Intel	CPU-intensive computations, model inference
Dsv5 Series	4	16	None	-	Intel	General ML data preparation, development
Fs_v2 Series	8	16	None	-	Intel	Compute-optimised workloads, parallel processing
Dsv5 Series	8	32	None	-	Intel	Large dataset processing, parallel computing
Easv4 Series	8	64	None	-	AMD	Memory-intensive analysis, large data in RAM
Fs_v2 Series	16	32	None	-	Intel	High-performance computing, CPU-intensive tasks
Dsv5 Series	16	64	None	-	Intel	Complex simulations, high-performance computing
GPU VMs - A10 Series						
A10_v5 Series	6	55	1/6 A10	4	AMD	Small ML model training, GPU-accelerated analytics
A10_v5 Series	12	110	1/3 A10	8	AMD	Medium neural networks, computer vision tasks
A10_v5 Series	18	220	1/2 A10	12	AMD	Deep learning training, image processing
A10_v5 Series	36	440	1 A10	24	AMD	Large model training, multi-modal AI research
A10_v5 Series	72	880	2 A10	48	AMD	Multi-GPU training, large-scale ML experiments
GPU VMs – H100 Series						
H100_v5 Series	40	320	1 H100 NVL	94	AMD	LLM training, cutting-edge AI research
H100_v5 Series	80	640	2 H100 NVL	188	AMD	Large LLM training, multi-GPU AI research