DATA CAPABILITY

GUIDANCE

UK Statistics Authority
Awdurdod
Ystadegau'r DU

Data Capability Guidance

Table of Contents

INTRODUCTION

Accreditation of processors under The Digital Economy Act 2017

The Research powers in the Digital Economy Act 2017 (DEA) facilitates the linking and sharing of datasets held by public authorities for accredited research in the public good. The Act provides a requirement that organisations wishing to become processors or obtain personally identifiable data and then link, match, or process these, must be accredited to ensure that their security environment, controls, capability, and processes are satisfactory to protect data.

Under the DEA the UK Statistics Authority(UKSA) is the statutory accreditor of processors, researchers and projects accessing and processing data under the Act. To oversee the processes used by the UKSA to accredit processors, researchers and projects, the National Statistician has established a Research Accreditation Panel (RAP).

The overall accreditation process is outlined in more detail on the UK Statistics Authority website. The current guidance supplements the existing guidance by providing additional information on what are expectations for the data capability controls of the framework and how these are assessed.

Our common commitment to transparency

This guidance is underpinned by our common commitment to transparency. Transparency for the accrediting body is based on clearly communicating our expectations from the accreditation exercise, all steps of the process by which we accredit data processors, and the maturity assessment against each control along with a set of improvement actions agreed with the processor. Our goal is equally clear, we want to enable safe and responsible research under the Digital Economy Act 2017 that meets and exceeds the needs of the research community.

At the same time, we envisage that processors are equally transparent against all control areas to realise the benefits presented by the Digital Economy Act. Transparency is the key enabler in overcoming barriers in data access, improving the reusability of data, data products and code across different Trusted Research Environments. As a large volume of high-quality data are a critical component of any research environment, removing such barriers offers a unique opportunity to attract high-impact research in the public interest. A consistent and transparent accreditation process encourages data suppliers to further invest in the digital economy with the assurance that their data is used safely, legally, and ethically.

Transparency is valued by the research community, who are asked to make the time-sensitive decision to identify the ideal environment offering not only data but a quality service to conduct research safely, ethically, and legally. Demonstrating openly the capability of processors to host projects allows researchers and research sponsoring organisations to make informed decisions without the uncertainties of an unknown environment.

Last but certainly not least, the public has an increased awareness and vigilance of how their data are used. The robust data capability framework underpinned by our commitment to transparency offers the public the much-needed assurance that their data is not misused and enhances the relationship of trust to statistics and statistical research.

Regardless of how strict or robust the proposed controls are, without transparency the digital economy is stagnating outside data silos, opportunities to collaborate without arbitrary barriers, and realise tangible benefits by delivering deeper insights into all aspects that matter to the public are missed. The data capability framework offers a robust set of scalable and measurable controls, through a set of transparent processes to enable data processors realise the benefits of the Digital Economy Act.

DATA CAPABILITY FRAMEWORKS

Control framework

The primary component of the accreditation framework is the control framework, comprising of a set of specific controls under each control area. The new data capability framework consists of a set of 30 controls separated in five key areas:

1. Research Governance
2. Data Governance
3. People capability
4. Service provision
5. Processor Accreditation obligations

Considering the inherent overlaps in controls in these areas we have determined some key questions to help you provide focused evidence against each control. Data processors would need to demonstrate their capability against all five areas to be accredited. The scope of the accreditation is limited to the controls that apply to the functions of the processors under the Digital Economy Act 2017 for the provision and/or preparation of data. Processes, functions, and systems not used under the Digital Economy Act will not be considered as evidence against controls and won't be assessed. Throughout this guide we will distinguish the controls for the provision of data, preparation of data and both using the following colour annotations.

COLOUR ANNOTATIONS FOR PROCESSOR FUNCTIONS

■ Provision of data

■ Preparation of data

■ Preparation and provision of data

In line with the control framework, we will set out what we expect from processors to meet this control. We won't delve into the controls and control areas yet as it is worth explaining the maturity assessment used to assess these controls.

Maturity assessment framework

The second component of the accreditation framework is the maturity assessment framework. Any processor going through the accreditation process, as well as the accreditation assessors, must have a clear view of how they the assessment works and applies against each control and throughout the control framework. Capability under this framework is seen as a dynamic process where organisations gradually mature. The benefits of a mature organisation are tangible. Mature organisations are capable of offering a better service to the research community and thus attract more high-impact and sponsored research projects. Additionally, they demonstrate their capability in curating data safely, legally, and ethically, and are trusted by data owners to curate more data. Finally, as maturity improves processors demonstrate their

trustworthiness to the accrediting body requiring less frequent auditing and accreditation review cycles.

The journey of an organisation toward maturity consists of six key areas of assessment:

1. Scope of controls: The processor controls cover all functions of the processor under the DEA.
2. Testing of controls: The processor maintains evidenced processes to test the effectiveness of controls (between accreditation reviews).
3. Assurance provided by evidence: The evidence provided are specific to the control and sufficient in detail.
4. Use of Management information: The processor has a set of MIs which are regularly monitored and used to inform decisions.
5. Naturally embedding of controls: The processor should demonstrate that staff understand the importance of the controls beyond compliance and strive to naturally embed the controls into their business-as-usual processes.
6. Approach to controls: Processors should strive to proactively explore ways improve their capability. Note that not all control areas necessitate a proactive approach.

Opinion	Controls (scope)	Controls (testing)	Evidence assurance	Use of MI	Embedded in culture	Approach
Mature	All	Consistent	Robust	Consistent	Yes	Proactive
Maturing	All	Ad-hoc	Good or Good-improving	Partial	Partial – improving	Mostly reactive
Capable	All	None	Capable or Good-stable	Partial	Partial – stable	Reactive
Partial	Most	None	Partial	None	No	Reactive
Minimal	Some	None	Sparse	None	No	Reactive

The result of the maturity assessment, expressed as an opinion falls under five categories. A processor must demonstrate a capable maturity level to be regarded as accredited under the Digital Economy Act. Ultimately, the level of maturity is an indication of how trusted these research environments are, in terms of the trust of the research community, the data owners, and the research sponsors. ***In line with the maturity assessment framework, we will set out what we anticipate as evidence against each control against the key areas of assessments.***

Opinion	Rating	Definition
Mature	🟢	Processor demonstrates a comprehensive set of controls that are tested and supported by consistently produced management information and robust evidence. Processes to regularly benchmark these controls are in place. Processor has adopted a proactive approach in improving its resilience and capability
Maturing	🟢	Processor demonstrates a comprehensive set of controls that are well documented and supported by management information and evidence. No improvement is required, but improvement actions have been recommended to further enhance the capability of the processor.
Capable	🔵	Processor demonstrates a comprehensive set of controls that are sufficiently documented and supported by evidence. Use of management information in consistently informing decisions can be improved. No improvement is required, but improvement actions have been identified to further develop the capability of the processor.
Partial	🔴	Processor demonstrates a comprehensive set of controls that are partially documented or not supported by sufficient evidence. Improvement is required.
Minimal	🔴	Processor demonstrates some controls that are partially documented, or ad-hoc supported by partial evidence. Significant improvement is required.

As any organisation can improve their maturity assessment, they can revert to a lower maturity assessment at the next round of accreditation. For instance, an organisation assured as good or maturing might slip back to capable at the next accreditation if they fail to address any improvement actions since last accredited.

Under the maturity assessment framework each control is assessed separately. The maturity of the data processor is calculated using a weighted algorithm considering all applicable controls. The algorithm weights put more emphasis on controls whose implementation provides more significant capability benefits and controls which if not implemented can lead to more higher risk exposure. The rounded average of all weighted controls produces the maturity opinion for each control area as well as the overall maturity opinion. This means that a processor might be regarded mature considering research governance even if not all controls are assessed as mature. Any optional controls that don't apply to a data processor are not assessed and this won't affect the maturity opinion.

For a processor to be accredited it is required that:
1. no control is assessed as partial or minimal, and
2. the weighted average of all controls is assessed as capable or above.

We encourage data processors to publish their maturity assessment opinion once accredited and this will be also published as part of the public register on the UK Statistics Authority website.

At this stage it is important to highlight that the maturity assessment is an opinion of the assessors on the capability of the data processors <u>for the functions that they offer</u>. This opinion should not be used to compare the different processors with each other without considering the functions and services they provide. There is no basis of comparison between a processor providing a limited scope of functions around the provision of data and assessed as mature and another processor providing a wider scope of functions around the provision and preparation of data and assessed as capable. For data processors, the main benefit of the maturity framework is to determine what best practice looks like for a specific control and service provided, and to inform the design and development of capabilities that best serve the research community.

Additional incentives for achieving maturity

The maturity assessment framework outlines a more thorough process, not limited to policies and procedures but how these are applied and evidenced in the processor's environment. Undoubtedly, the new framework requests processors to provide more robust and consistent evidence regarding their capability to be assessed at a higher maturity. In order to compensate this, we have adjusted the data capability framework to provide clear incentives to such processors.

While processors that have demonstrated sufficient assurance of their capability will be audited on an annual basis as before, accredited processors at the maturing and mature state would be audited less frequently. They would also need to provide evidence on a smaller sample of data capability controls. Consequently, the investment to improve the capability and quality of service offered to the research community, as reflected in the maturity assessment, is recognised by the accrediting body. This reinforces the

ACCREDITATION REVIEW FREQUENCY AND SCOPE

		Frequency	Regular audit scope	Ad-hoc audits
Mature	🟢	Every three years	Sample of controls (excl. mature controls) and improvement actions	1-3 audits
Maturing	🟢	Every two years		2-4 audits
Capable	🔵	Annually	All of controls and improvement actions	2-4 audits

relationship of trust between the accredited processor and the accrediting body.

For processors with a good / maturing assessment it is important that processors can demonstrate not only how the assessed sample controls is showing improvement, but specifically how the improvement actions and recommendations issued by the assessor have been addressed. If no improvement is shown, good/maturing controls can revert to capable.

In addition to regular scheduled audits, the accrediting body would have the right to conduct ad hoc audits. This is to verify that controls remain relevant and robust in demonstrating the capability of the processor. This is particularly important when

systems, process and data and people capability change or when controls are marginally assessed as capable.

Regardless of the maturity level each organisation would need go through the accreditation process, reviewing all controls, every five years. For instance, a mature organisation first accredited in 2022, would go through an accreditation review in 2025, followed by a full accreditation review 2027. A schedule of accreditation exercises depending on the accreditation assessment outcomes is presented in Annex B.

Controls and control areas

Research Governance
How we enable responsible research

Research governance is concerned with the ongoing management of accredited researchers and projects under the Digital Economy Act. The controls ensure that the processor has the capacity to manage users in their environment for specific projects, using appropriate data under recognised legal and ethical frameworks.

> **C.1.1.** Maintain sufficiently detailed records, including any accreditation conditions, of all projects in the processor's environment.

What we expect: A project catalogue in the environment with a sensible amount of information, including specific accreditation conditions, recorded consistently. The processor can decide the amount of the information and the format of this catalogue. There is a reasonable expectation that essential fields, as determined by the accrediting body (Appendix A-Part E), to manage projects are recorded. All processor involved in the provision of data must provide evidence against this control.

Depending on the level of maturity of the processor we anticipate:
- Improved quality and usability of the project catalogue enabling support staff to make clear decisions.
- Management information should be in place to examine the characteristics of projects in the environment.
- Processes to quality assure and review this catalogue, evidence of audits and improvement actions, metadata scheme for the catalogue identifying clear links between projects, researchers and specific data views should also be in place for a mature processor.
- Proactiveness is defined via the use of information from the project catalogue to inform decisions on service delivery. For instance, use of MI on the characteristics of projects to determine what data to acquire.

> **C.1.2.** Monitor research taking place in the processor's environment, identify and report any significant deviation from accredited project conditions.

What we expect: This control appears simple but is in fact rather complex in nature. We expect that the processor has
- a process to check if projects, <u>not only statistical outputs</u>, are within scope and escalate internally and externally, as and if necessary,
- evidence of sample audits on projects including scope checks,
- guidance and/or training to researchers on what constitutes research out of scope and how it is handled,
- guidance and/or training to support staff on what constitutes research out of scope and how it is handled,

- records of incidents related to research out of scope, as well as requests for scope change to the accrediting body, along with the decision of the accrediting body.

Depending on the level of maturity, we anticipate:
- Consistent and easy to understand information to determine if a project is within scope are easily accessible to support staff. For instance, support staff can easily access information to understand the objectives of a project, the conditions of accreditation, the specific datasets involved, and any restrictions regarding these datasets.
- Staff are confident in determining if a project is out of scope and understand their responsibility and the exact steps to report to the service and the accrediting body a scope violation. The decision of the accrediting body, if required, is clearly recorded against each request to change scope.
- Management information on incidents include analysis and review of incidents related to out-of-scope research.
- A schedule and an audit plan of sample project audits is in place.
- Guidance and training provided to support staff and researchers is regularly reviewed and feedback received by those users on guidance is actioned.
- Managers of support staff assess the competency of support staff following the process and understanding the guidance provided. Equally, support staff foster an open culture encouraging researchers to discuss potential scope changes. This culture is supported by guidance and training.
- A proactive organisation improves the capability of identifying projects likely to move out of scope, using incident and service information to inform engagement and audit activities.

C.1.3. Clearly communicate the available statistical/analytical software in the environment and manage any changes to software and its impact to research.

What to expect: Information on the statistical and analytical software available in the environment. For software depending on additional modules (e.g., R packages, Python libraries, SPSS Add-on modules, MATLAB toolboxes) researchers should be aware of what is available in the environment, and the process of requesting additional software. When changes are to be made of software, impact assessments are conducted by the processor, shared, and communicated clearly with the affected researchers.

Depending on the level of maturity, we anticipate that:
- There is evidence of a regular review of the publicly available information on software. Feedback from researchers on this information, as well as software issues and suggestions are is captured and feeds into decision making.
- The processor can easily identify all software instances including modules, for each project.
- Metrics on software utilisation are in place and used to inform relevant decisions and impact assessments.

- When changes to software are required, there are established processes for testing and deploying, responding to researcher queries, and rolling back to previous versions if necessary.
- When changes to software are required, impact assessments are undertaken, and communicated to researchers whose research might be impacted.
- A mature processor is expected to proactively explore new software and technologies to be used in the environment and engages with the research community to identify emerging opportunities and trends.

C.1.4. Evidenced processes for managing accredited researchers in the processor's environment throughout the researcher journey.

What to expect: The key expectation here is that a set of procedures is in place to manage the accreditation (first accreditation and renewal), training, allocation of researchers to projects, and revoke or suspend of researcher access from projects. This must include fully accredited and provisionally accredited researchers, and all other external users that might access data or statistical outputs in the environment (e.g., peer-reviewers and research supervisors under the approved researcher scheme). These procedures need to be evidenced in consistent researcher records supported by any additional evidence as specified by the internal procedures in the processor's environment.

A mature data processor is expected to maintain and review the following processes to cover the entirety of the data journey:
- researcher accreditation
  includes any process the processor put in place to administer researchers in their systems, check their accreditation status, notify them of accreditation status changes and renew their accreditation.
- project application and accreditation
  includes any processes the processor put in place to administer projects in their systems, capture the application and data ethics forms, notify researchers of progress of their application and accreditation status changes, record any project-specific restrictions, and manage project change requests.
- training,
  includes any processes the processors put in place to manage training courses, administer researchers requiring to be trained or retrained.
- access to data,
  includes any process the processor has in place to manage researchers' access to data, enable, revoke, or suspend access to data.
- analysis within the processors' environment,
  includes adding or removing researchers to/from projects
- dissemination of outputs,
  includes inviting peer-reviewers or supervisors to access data and interacting with the support team to apply SDC on outputs.
- end of project activities
  includes procedures to revoke access to project data, access archived data and reusing code and data for other projects.

Non-compliance incidents require flagging both the researcher and the organisation affiliated with the researchers (e.g., academic institution, commercial sector organisation, government department). It is important that management information can be translated from a researcher to an organisation level. For instance, any processor should be able to determine the volume of researchers and incidents by a particular organisation in the environment.

We expect processors to:

- Maintain a wide set of procedures regularly reviewed and updated. There is a clear link between the procedures and the expected evidence of their implementation. For instance, when reminders are sent out to a researcher to renew their accreditation, this is clearly recorded, alternative routes of notification exist, the notification schedule is sensible, and notification limits are adhered to.
- All users of the service are accounted for, along with their accreditation status, any incidents and near misses, and any access restrictions.
- Record all incidents of non-compliance against procedures and review these incidents regularly.
- Produce and review management information on the management of procedures and incidents related to procedures. There is evidence that this information is used to inform decisions on service delivery.
- The processor has mechanisms to collect and analyse feedback from researchers on their experience throughout the researcher journey.
- User management information is consistent and easy to use for all functions that interface with researchers. For instance, when managing access for a project, the support team can easily access the appropriate amount of information to make decisions. A mature organisation limits that information to what is necessary for a specific function.
- Staff are aware how procedures operate, the escalation and contact points for each procedure and staff feedback is considered when reviewing these procedures. Procedures are kept centrally and are easy to understand and implement.
- A mature organisation is proactive in scaling up controls and management information as the volume of researchers increases, investigate when new user profiles are emerging, act on user feedback and regularly benchmark processes.

C.1.5 Ethical frameworks in place for all DEA accredited projects and all ethics processes in the processor's environment remain transparent and auditable.

What to expect: A key requirement of the Digital Economy Act 2017 is the need to uphold high ethical standards. This does not mean that we expect every processor to establish their own ethics committee. However, there is a reasonable expectation that there is a structured, evidenced, and transparent process to assess the ethical risk of research projects and manage the mitigation of that risk. In summary, we expect that ethical aspects are considered for all projects, any ethics concerns are recorded and any restrictions arising from these concerns are adhered to.

Depending on the level of maturity, we anticipate that:

- A procedure to consider ethics against each project is established and there is clear evidence in the form of ethical assessments.
- Ethical assessments clearly specify if a project is ethically safe to proceed or if specific restrictions apply to ensure it adheres to high ethical standards.
- Staff can easily find these assessments and any restrictions and along with other project information can effectively assess whether ongoing research remains within these restrictions. For instance,
  - if a set of variables poses a significant ethical risk for a project staff can easily ascertain that these variables are not included in the data made available to researchers.
  - if a particular output format is not ethically acceptable (e.g., inferring causality when statistics only infer correlation) this is not released from the environment.
- A mature organisation can produce qualitative and quantitative metrics to assess the ethical risk of research within the environment.
- Staff have sufficient guidance, support, and training to understand how to implement ethical restrictions on statistical research and statistical outputs and are confident when communicating ethics to researchers. Evidence that guidance, training, and support mechanisms is reviewed, feedback from staff is used to improve these artefacts.
- Managers of support staff assess the competency of support staff following the process and understanding the guidance provided.
- A proactive processor can use metrics and qualitative observations to detect emerging ethical risks, changes in the risk profile of projects and inform more robust ethical frameworks.

C.1.6. Record and review any requests for specialist software for all projects. Optional

What to expect: As researchers from different disciplines join the processor's research environment, they might have particular requirements for software or software modules. The processor should be capable of assessing the needs of the research community and improve the reusability of code and software in their environment. Given that software is a tool, its use can present both benefits and risks. For instance, software and code can be used to infer information from pseudo-anonymised data or be misused (e.g., using an ethnicity estimator with disregard to its terms and conditions of use and present ethnicity proxies without the estimator probability). A review of each request must take place when enabling specialist software in terms of security and the use of the software license. In terms of capability, there is no need to review each request but regularly review a larger number of requests to identify user needs and patterns.

Although this control does not pertain to code brought in by researchers, but only analytical and statistical software and software modules, mature organisations would be able to extend the recording of all code brought it the environment and encourage code reuse. For instance, code developed by a researcher to clean a specific dataset

or standardise specific variables in that dataset could be catalogued and made available to multiple projects using this dataset.
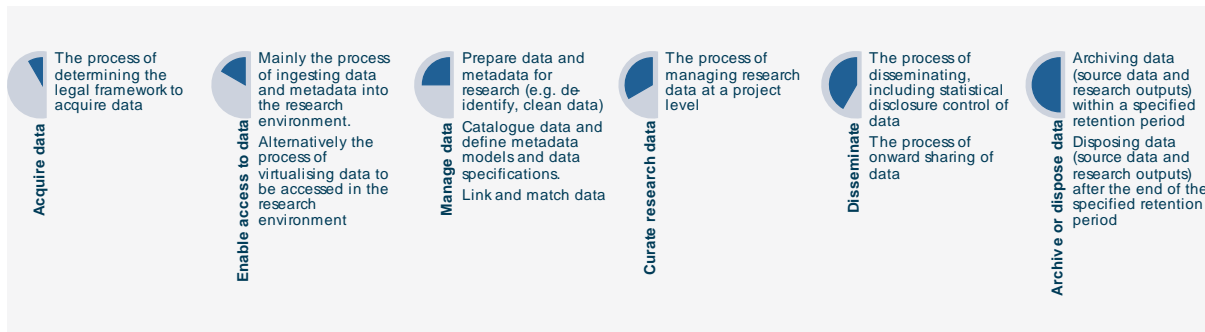
Depending on the level of maturity we expect that:

- Requests for software are recorded along with the researcher need for this. When software is requested the terms and conditions of service are considered in relation to the accreditation conditions and any ethical restrictions.
- Researchers are provided with enough information on how they can request new software and ingestion of software modules in the environment.
- Researchers might have identified software or software modules that can be regularly used with specific data in the environment. There is an untapped potential for mature processors to recognise these patterns and have a process to make such tools more widely available to researchers.
- Management information is produced per researcher, project, affiliated research organisation and data to inform decisions on software to be made available and software modules to be used.
- Staff are confident and supported when making decisions regarding software requests and can easily access relevant information on the project accreditation and ethical approval.
- The value of this control lies in enabling proactive decisions and promoting code reuse. A proactive processor has processes to use management information and user feedback to determine a process to reach decisions on new software and reusing code and assess the impact of these decisions. The decision-making process should be properly evidenced and guidance and processes for enabling the re-use of code will be in place.

Data Governance
How we curate data

Data governance relates to the ongoing management of data, metadata, and code. The controls ensure that the processor is capable of curating data and code safely, legally, and efficiently throughout the data lifecycle. Processors must demonstrate that appropriate controls are in place for each stage of the data lifecycle. The latter can be defined differently in different processors, depending on the functions they offer, but largely includes the following stages.

Acquire data	**Enable access to data**	**Manage data**	**Curate research data**	**Disseminate**	**Archive or dispose data**
The process of determining the legal framework to acquire data	Mainly the process of ingesting data and metadata into the research environment. Alternatively the process of virtualising data to be accessed in the research environment	Prepare data and metadata for research (e.g. de-identify, clean data) Catalogue data and define metadata models and data specifications. Link and match data	The process of managing research data at a project level	The process of disseminating, including statistical disclosure control of data The process of onward sharing of data	Archiving data (source data and research outputs) within a specified retention period Disposing data (source data and research outputs) after the end of the specified retention period

Although data governance includes the largest number controls, providing assurance against those should be straightforward.

> C.2.1. Maintain clear and consistent records of legal agreements that outline how data is accessed, processed, and used in the environment.

What to expect: We expect that the processor is capable of linking each and every dataset to the relevant legal documents outlying the conditions on how the dataset can be processed, used, disseminated(e.g., statistical disclosure control thresholds and onward sharing of data), retained and disposed. A mature processor is capable to easily translate these documents into ready-to-use information.

Depending on the level of maturity we expect that:
- The processor holds central and up-to-date records of data linked to the relevant legal documents. Processes exist to review and audit this repository and the relevant documents.
- An audit and review schedule for these legal documents is maintained and evidence that outcomes of these audits and reviews are actioned are provided.
- A process exists and specific incident record are captured when the conditions set in the legal agreement are not fully adhered by staff or research users. The process should ensure that data owners remain informed throughout the incident investigation and satisfied with the remedial course of action.
- Staff in the processor's environment know where to access the legal agreement and confident that they can understand which legal gateway is used, as well the restrictions imposed on the data.

- A mature organisation is able to consistently extract information from legal documents and effectively translate these into accessible and comprehensive information (e.g., database record format)
- Management information on the legal gateways used, the restrictions imposed by different data owners is produced and used to inform decisions. Improvement actions, as result of reviews are tracked and actioned.
- A proactive organisation is capable to use a variety of management information to influence future data acquisitions and data access arrangements, with the aim of removing unnecessary restrictions on data.

C.2.2. Evidenced and appropriate procedures to manage data, metadata, and code in the environment.

What to expect: This is a broad control and processors should break this down for data and metadata, and code. Processors ought to examine the former for each stage of the data journey that applies to them. When processes are automated, there is sufficient human oversight of these processes to ensure that they operate as expected.

In more detail, we expect that processors maintain procedures on how:
1. data and metadata are acquired
2. data, metadata, and code are ingested physically or virtually in the environment
3. data is accessed by all users
4. data is linked and matched
5. data is prepared (e.g., cleaned, de-identified, sliced)
6. metadata is standardised in the environment
7. data, metadata, and code are curated for each project
8. research outputs are appropriately controlled for statistical disclosure
9. research outputs can be released from the environment
10. research outputs can be accessed by non-accredited users (e.g., project supervisors, programme managers and peer reviewers) within that environment
11. data can be shared to another organisation on behalf of the data owner
12. data retention is reviewed
13. data is archived and disposed

Depending on whether the accredited processor is involved in the provision and/or preparation of data not of these procedures would apply. Holding these procedures is not enough for an organisation to be accredited. What we anticipate, depending on the maturity of the processor is that:
- All procedures remain up-to-date and clearly specify how the processor is expected to demonstrate compliance.
- An audit schedule is maintained, and evidence of audits is provided against each procedure. Actions resulting from audits are actioned. Conducting audits is a requirement when automated procedures are implemented.
- Staff delivering specific functions of the processor understand how procedures are implemented and the expectations of how procedures are

evidenced. Feedback from staff is used to inform how procedures are developed.
- Management information on incidents or near misses relating to non-compliance to procedures are consistently reported and actioned.
- When improvement actions are identified against specific procedures, these are clearly owned and actioned.
- Guidance, training, and support is provided to staff, and is regularly updated and reviewed. Staff understand what these procedures aim to achieve and can challenge their implementation.
- A mature organisation proactively reviews the implementation of procedures, using a variety of management information and feedback by researchers and support staff, issues and implements improvement actions.

C.2.3. Clear procedures and records managing data and code brought in by researchers.

Optional

What to expect: When researchers have the option to bring in data or code they hold into the research environment, the processor must be capable of determining what is allowed within the scope of the project accreditation and how to verify that these can be legally accessed in the environment. For instance, it could be preferable if the data processor could acquire open or safeguard data instead of receiving the same dataset from a researcher. In any case, there is a reasonable expectation that all data and code brought in by researchers are accounted for. For mature organisation this offers valuable insights on what data researchers require and can inform future data acquisitions and what code researchers require to inform the availability of software and code in the environment.

Depending on the maturity of the processor we expect that:
- Processors maintain a log of all data brought in by researchers with sufficient metadata, including the legal gateway.
- Processors set a minimum set of data specification requirements for researcher data to be ingested.
- Processors maintain a log of all code brought in by researchers with sufficient description and metadata to understand what it is trying to achieve.
- Processors have a procedure outlining how researchers can bring in their own data and code, the researchers' obligations and legal, ethical and project accreditation aspects.
- Staff supporting the processors environment must be able to easily access sufficient information about the project and support by subject matter experts, when required, to inform their decisions when enabling access to research data and code in the environment.
- Management information is produced and reported on the data and code that is brought in by researchers. This information is used to inform decision regarding the service.
- Staff are aware of the risks of third-party data and code accessed in the environment and incidents of researcher data erroneously admitted into the researcher environment are investigated.

- A proactive processor is expected to adopt a schedule of regular reviews of researcher data and code, develop easy-to-implement standards (e.g., data specifications and code standards) and uses a variety of management information and feedback from staff and researchers to inform decisions.

C.2.4. Policies and procedures in place and tested to safeguard the confidentiality of data subjects in outputs.

What to expect: In addition to procedures and processes covered in C.2.1 this control sets specific expectations for Statistical Disclosure Control(SDC). The focus is the capacity of the processor to ensure the successful implementation SDCs.

Depending on the maturity of the processor, we expect that:
- The processor has a clear procedure in applying SDC against all data, with specific information and guidance for specific datasets and statistical outputs.
- Staff are able to easily access sufficient information on SDC rules that apply to specific datasets.
- When data is linked there is a reasonable expectation an agreement is reached between the different data owners regarding the appropriate SDC rules that would apply to the linked dataset in order to assure that the confidentiality of data subjects is safeguarded.
- When exceptions to the normal implementation of SDC rules exist, the processor maintains a clear record as well as the justification in line with any conditions stated of the project accreditation. Exceptions are reviewed regularly and incorporated into guidance and procedure documents.
- Incidents related to the disclosure control are recorded and investigated separately. Evidence of incidents reviewed, and lessons identified are expected.
- Management information are produced and reviewed relating to the application of SDC on data, as well as incidents related to disclosure control.
- Staff understand why SDC are important and their responsibility in implementing SDC in statistical outputs. They can easily access information (e.g., SDC rules) and have sufficient guidance, training, and support. Feedback on the process of applying SDC is captured and actioned. When staff do not comply with relevant procedures, this is recorded and appropriately managed.
- A proactive processor would explore the implementation of SDC in new and complex data (e.g., integrated data), produce and evaluate guidance on SDC.

C.2.5. Policies and procedures in place and tested to ensure reasonable deidentification of data.

What to expect: In addition to procedures and processes covered in C.2.1, this control sets specific expectations for the process of de-identifying data. This control applies to processor involved in the preparation of data and aims to assess their capability to deidentify data following robust statistical methodology and in line with legal, ethical and security considerations. In line with the Digital Economy Act 2017 data processors

must process data before these are disclosed, to ensure that it is unlikely that the person's identity could be deduced from the information (whether by itself or taken together with other information). The relevant safeguards in place must go beyond removing identifiers and take into account the wider data and information available to researchers, inside and outside the trusted research environment. For processors involved only in the preparation of data, there is an expectation of reasonable de-identification after the data has been processed.

Depending on the maturity of the data processor, we expect that:

- The processor has a clear policy, supported by procedures and guidance, on how data must be de-identified. After data has been de-identified, the processor ought to have a process to quality assure the data which must include a clear and informed assurance statement that data has been reasonably de-identified.
- Any deviations from standard practice or dataset-specific rules are recorded and regularly reviewed. The result of these reviews is incorporated into guidance provided to staff.
- De-identified versions of data are regarded as separate instances and managed accordingly (including separate data sensitivity assessments). The lineage between the original dataset and the de-identified instance is clearly evidenced.
- If researchers deem that a dataset has not been reasonably de-identified, the processor maintains clear procedures to investigate (which might include suspension of access to the data until the investigation is complete). All investigations must be recorded and reviewed regularly. It is clearly communicated to researchers that they need to report instances where they deem that data has not been reasonably de-identified.
- Staff understand the need to safeguard the confidentiality of data during analysis by providing and curating de-identified data. They are supported by guidance and training which is regularly reviewed, including feedback from staff. When staff do not comply with relevant procedures, this is recorded and appropriately managed.
- Management information is produced to assess the effectiveness of data de-identification as well as incidents and near misses. This information is regularly reviewed and used to inform decisions that affect the data preparation.
- A proactive processor performs sample audits on datasets, with focus on data which pose a greater risk of re-identification. They would evaluate guidance and processes and develop forward looking guidance for new datasets using new methods (e.g., synthetic data, privacy enhancing techniques).

C.2.6. Policies and evidenced procedures for the retention of all different data instances.

What to expect: In addition to general retention considerations, this control puts additional emphasis on the retention of all data instances for a dataset. This includes data extracts, de-identified, prepared, research and source data, and statistical

outputs. The main aim is that the retention of all types of data is governed consistently, and no data is retained longer than necessary.

We expect that processors demonstrate that:

- Retention plans, schedules and policies are in place addressing the retention of all data instances within the research environment. These are regularly reviewed and updated requiring that processors justify the ongoing retention of data. All processes related to the processing and provision of data include a retention review element.
- Retention reviews are regularly conducted, and outcomes of these reviews are translated into actions with clear ownership, completion dates and resolution. The audits are not concerned only with data available to researchers but sufficiently address how all data instances are kept.
- Staff understand the legal, ethical and security implications of retaining data and are aware of how retention artefacts (plans, schedules, procedures, and policy) can be used to achieve this. There is clear responsibility for specific staff members to coordinate retention reviews while all staff members curating data are held responsible for identifying and addressing gaps in retention plans.
- Management information on the retention of data is produced along with data utilisation metrics to justify the ongoing retention of datasets. This information is regularly recorded and reviewed in order to feed into decisions.
- Processors adopting a more proactive approach have clearly determined how dataset lineage should affect data retention, perform regular reviews to refine this process. In order to facilitate these reviews, the systems used are able to quickly extract relevant MI without the need to search into documents.

C.2.7. Clear records of all data in the environment. This includes open data, geography lookups and data extracts.

What to expect: This controls ensures that all data are accounted for and aims to improve the reusability of data and ensure the production of transparent data utilisation metrics. Of particular interest are data extracts (slices, views, or parts of a larger dataset). The control sets a clear requirement that these are sufficiently recorded which would entail that some, but not all, information is passed from the larger dataset into the data extract record. For instance, in some cases, data sensitivity assessments as well as legal gateway information would be shared between the parent record (original dataset) and the child record (extract of this dataset). In cases where data is further processed (variables on protected characteristics or potentially person identifiable fields are removed) a new assessment of data sensitivity is required.

Depending on the maturity of the organisation, we anticipate that:

- Data registers, catalogues or databases clearly identify supporting data (open and geography data) as well as data extracts.
- Procedures are in place and regularly reviewed and updated, to outline the relationship of parent-child data and the requirements in terms of the Data Protection Impact assessments, Data sensitivity assessments and information risk management records. When terms and conditions apply (e.g.,

safeguarded data) further work is done to ensure that data extracts and use of data is in line with these conditions.

- Information in data registers is easy to understand and in a format that enables timely and efficient analysis.
- Staff curating data are responsible for accurately recording datasets and performing quality reviews on the data register records. Guidance, training, and support is provided to staff and feedback is regularly captured to ensure that they remain adequately supported.
- Management information is produced to capture the relationships between datasets and assess quality information of the different records.
- Proactive organisations are able to trace the use of different data instances across their data estate and make informed decisions using on management information regarding lineage and data utilisation. This might involve negotiating for more specific data to meet the needs of the research community, refine retention reviews and inform the development of data products (e.g., complex lookups).

> C.2.8. The processor clearly specifies data ownership in their environment, including the ownership of research outputs.

What to expect: The control mandates that at all times the processor can clearly specify who the data owner, and the data controller for personal data, is for all datasets in the environment. The processor also specifies information governance arrangements for data, metadata, and research data. For organisations that prepare data, it is important that the ownership of processed data is defined. For organisations that provision data, they need to define whether data owners have implicit rights on research outputs allowing them to prevent the dissemination of research data, or if researchers solely own their research outputs. These arrangements should be made clear to data owners and data controllers as well as researchers.

Depending on the maturity of the data processors, we anticipate that:
- The processor has clear information governance arrangements for all data, which specify the following roles as a minimum;
    - data owner or data controller for personal data. Person(s) who own the data.
    - data steward/manager. Person(s) who curate data and metadata in the processor's environment.
    - data expert. Person(s) in the data owner's organisation who can respond to any queries regarding the data.
- The processor has a clear policy specifying the ownership of data in the processor's environment. This policy is regularly reviewed and communicated clearly to:
    - data owners before they enable access to their data in the processor's environment (at the acquisition of data), and
    - researchers during their accreditation and before they apply for an accredited project in the processor's environment.
- When conflicts arise regarding the ownership of data, these are well documented and used to produce lessons learnt.

- Management information include metrics relating to the ownership of the data and these are clearly used to inform decisions.
- Proactive organisations ensure that information governance arrangements are agile to enable timely decision making without introducing unnecessary delays and barriers in the flow of data under the Digital Economy Act. This can be enabled by increased availability and usability of governance information, contacts to data owners and standardised or streamlined communications.

C.2.9. Onward sharing of data by the processor is audited and recorded. Optional

What to expect: When the processor decides to onward share data to a third party on behalf of the data owner/data controller a clear audit trail is produced. This is used to inform decisions on onward sharing and disclosure as well as improving the overall service offered by the processor. This includes when data is been prepared by an accredited processor (for preparation of data) in order to be onward shared to another environment (for the provision of data) under the Digital Economy Act 2017.

We anticipate that:
- Clear records are in place to record the onward sharing of data, including the legal gateway, the purpose, the expected use, and benefits from onward sharing.
- Processes are in place to review these records in terms of quality and consistency. These reviews are regular, and outcomes are recorded and translated into actions with clear ownership and timelines.
- Processes and legal agreements are in place in case data is shared in error with clear instructions of how these must be disposed, and their disposal verified. Such incidents are recorded as a separate category, enough information is provided on their investigation and resolution.
- Guidance, training, and support is provided to staff involved in the onward sharing of data. These are regularly reviewed and take into account feedback from staff, and parties involved in the data sharing. Staff can easily access information necessary to inform decisions regarded to data sharing and clear responsibilities are set to verify that all data shares are done within the appropriate legal gateway and present a clear public benefit.
- Management information on onward disclosure is used to inform decisions that improve the service offered by the processor. This includes qualitative information.
- Proactive organisations aim to automate the production of the audit trail for onward sharing, detect abnormal patterns of data sharing and use management information to inform wider data management and curation strategies in the processor's environment.

C.2.10 Auditable systems and practices to ensure that access to proportionate data provided (data minimisation)

What to expect: In most cases researchers won't need access to all instances of the data and must respect the data minimisation criterion under the UK GDPR. For instance, for a project exploring data in the last decade, data from the 1980s are irrelevant and must not be made available. In some cases, researchers should not have access to specific variables, for instance, when including variable pertaining to a protected characteristic poses a significant ethical risk which cannot be mitigated. The same applies to support staff who should be able to access administrative information within the scope of their role.

This simple control ensures that, depending on the maturity level:
- The processor can provide a detailed view of the data made available for each project. The process can easily demonstrate that the data is appropriate and necessary to deliver the research objectives, respects the conditions of accreditation and any ethical restrictions.
- The processing environment has conducted a Data Protection Impact Assessment, has specified clear roles for support staff and adjusted access to data appropriately to each role e.g., on a need-to-know basis
- The processor has processes that enable them to efficiently interrogate data for each project, investigate discrepancies and take remedial action when required. Audit reports are drafted for all sample audits and remedial actions are tracked and reported.
- An audit schedule is in place to regularly assess projects and access. For mature organisations, audits consider the sensitivity of the data involved and the ethical risk of the research project.
- Management information is produced to explore the utilisation of data in the environment at a lower level of granularity (e.g., variable categories would suffice). This information can inform data specifications and must inform retention to justify the current need for curating specific datasets or variables.
- Staff and researchers understand why data minimisation is important, are aware of their responsibility to report cases where this principle is not upheld, and relevant guidance and training are in place and reviewed regularly.
- A mature organisation will be proactive to identify how optimised data utilisation can inform new acquisitions, production of linked and integrated datasets, availability of research-focused data products. A mature organisation is proactive in exploring how data minimisation can be implemented more effectively, in terms of technologies and practices in managing data access.

People Capability
How we empower staff and researchers

This control area is concerned in how processors can demonstrate that both researchers and staff are competent, skilled, experience and supported to operate in the processor's environment. Unlike the rest of the control area this consists of a single but complex control. Processors would need to verify three key criteria:
1. competency of staff (relevant skills and/or experience)
2. security clearance
3. training and support of staff and researchers.

These three criteria are to be explored against <u>all functions</u> of the processor. Practically this means that a processor must be able to list all functions they offer under the Digital Economy Act and provide evidence on how these three criteria are satisfied. Naturally when controls overlap the processor can associate the same evidence with different functions. If any of these functions are to be undertaken by researchers (matching, linking, statistical disclosure control) the process must verify that researchers also meet these criteria.

This includes, but is not limited to, functions as such:
• Design and delivery of training courses
• Data preparation
• Manage/curate data
• Data de-identification
• Matching and linking data
• Applying statistical disclosure controls on outputs
• Providing advice on projects, software, and code.
• Respond to requests and queries by researchers and users.
• Onward sharing of data

C.3.1. Demonstrate that staff have the relevant skills and/or experience, security clearance, training, and support to provision all functions of the processor in line with relevant policies. Evidence of ongoing assessment and development of staff.

What to expect: This control requires that the roles and responsibilities of staff in the processor's environment are clearly set. It is up to the processor to define how they wish to assess the competency of staff, but there is a requirement that they clearly evidence how staff are developed and supported. When staff are deemed not to comply with relevant policies or following established procedures, there is an expectation that appropriate remedial action is undertaken, considering the responsibility of the processor in safeguarding the data trusted in their environment.

We expect that:
• Processors define how the skills and experience of staff is established during the recruitment process and evidence that it is been assessed as part of the performance review.

- Guidance is easily accessible to staff and when reviewed, feedback from staff is considered. Staff are aware of where to request support and training available to them.
- Performance reviews of staff are regularly conducted and consistently recorded. Any actions, as additional training, is clearly captured and timelines are set.
- Data or functions requiring additional security clearance is clearly recorded as such and staff are aware of these restrictions. The evidence of the security implementation of this control, as security and access audits, are captured under the security controls.
- Incidents related to staff behaviours, including human error and non-compliance to policies, are recorded separately. Incident information is used to inform performance reviews, staff management and training. Following an incident involving human error, proportionate action is taken to safeguard the safety of data and confidentiality of data subjects.
- Incidents related to researchers' behaviours, including human error, malicious attack, non-compliance to policies, are recorded separately. Depending on the nature and impact of the incident, proportionate action is taken which might include suspending or withdrawing of researcher accreditation, or re-training researchers.
- Management information is regularly produced and reviewed to reflect on the capability of staff to support the functions provided. Metrics are regularly used to inform decisions regarding staff management.
- Proactive organisations have forward looking plans to develop their staff and trace staff development using quantitative and qualitative metrics They are also able to evaluate the effect of improving staff capability in service delivery and manage staff workload.

Service provision
How we interact with researchers

Ultimately, the main purpose of curating data in the processors environment under the DEA is to offer a high-quality service that empowers researchers to conduct impactful research in the public interest. This control area is concerned with that interface between the users of the service and the service providers. All controls aim to ascertain that processors are transparent and inclusive on the service they provide so that researchers can make informed decisions about their project.

C.4.1. Publicly provide information the functions and services provided, including accessibility and performance information regarding these services.

What to expect: The processor should list all services and functions provided as well as key performance metrics, as determined by the accrediting body (Appendix A-Part D), about these services.

We anticipate that processors:
- Make publicly available a set of services and key functions they provide to researchers. This must include key performance metrics so that researchers can make informed decisions about their project.
- The information provided is clear to understand and metrics are produced in line with the requirements set by the accrediting body.
- Information must be made available to researchers to contact the accrediting body directly if they believe that the information on the services and functions as well as performance is not accurate.
- Performance information is provided in a timely manner, as agreed with the accrediting body, and processors used it along with other metrics to inform decisions on service provision.
- The production of this information is embedded in the regular reporting streams of the service and internal processes are in place and reviewed to consistently produce and use metrics.
- Performance information is reviewed regularly to explore trends and patterns and continue to improve the service offering. Outcomes based on performance information are translated into specific actions with clear ownership and timelines. The impact of change requests and improvements is assessed against these performance metrics.
- In addition to performance information, a mature organisation would be able to communicate information on the accessibility of the functions and services. For instance, if access to a particular dataset or software is limited to specific researchers or service fees are required to link datasets, it is important that this is clearly communicated to researchers as it can potentially affect the level of service provided.

C.4.2. Provide sufficient information on how researchers can interact with the service before researchers access the environment.

What to expect: In addition to a list of services and functions, there is a reasonable expectation that each processor provides sufficient information on how researchers can interact with the service throughout the researcher journey. The latter consists of the following key stages:

- researcher and project accreditation,
- training,
- project application,
- access to data (including bringing researcher and open data in the environment, communicating data restrictions, quality, and assumptions)
- analysis within the processors' environment (including how project data and code in the environment is managed, importing code and software into the environment, and the use of software tools and code repositories)
- dissemination of outputs (including how to apply SDC on different data, extracting statistical outputs, data, and code from the environment), and
- end of project activities (includes information of what happens when a project is closed upon completion, how research data will be retained, specify the ownership and intellectual property of research data, capability to reuse outputs, data products and code for other projects.

Although the detailed researcher journey can vary in different processing environments it would be useful to communicate this journey consistently to the research community to support a common understanding of the data access process across the research community. There is a reasonable expectation that processors are able to present the researcher journey based on the key stages presented above.

In addition to the researcher journey, researchers should be made aware of key contacts within the processor's environment including staff providing support with:

- research queries,
- project application,
- project accreditation,
- data ethics,
- data acquisition,
- data security,
- methodology, code, and software,
- data experts, and
- statistical disclosure control.

We anticipate that processors should demonstrate the following:

- The research journey in the processor's environment, along with timelines for each stage is made clear to researchers before accessing the environment. Sufficient and accessible information is provided at each stage.
- Contacts are easily accessible outside or within the processor's environment depending on the stage of the researcher journey.

- Mechanisms are available to researchers to offer feedback to processors or the accrediting body directly. Evidence that feedback is reviewed and considered in making service decisions.
- Information and material related to how researchers interact with the service is regularly reviewed, considering feedback from the wider research community. As organisations mature, we expect the quality of the information provided to improve.
- Proactive organisations, actively engage with the research community to capture their expectations on the material publicly and internally available and issue clear actions to improve their offering.

C.4.3. Effective, accessible, inclusive, and auditable systems to respond to requests by researchers.

What to expect: The processor has an obligation to maintain efficient mechanisms to respond to requests by researchers. These must be also inclusive and accessible. Information should be consistently recorded, quality assured and reviewed to inform decisions on service provision.

Depending on the maturity of the data processor, we anticipate that:
- The processor maintains a platform that allows researchers to submit queries and requests and for support staff to respond and trace how these are resolved.
- As queries and requests can relate to various parts of the researcher journey, any platform deployed by the processor must be inclusive and accessible. Anyone interacting with the service and deeming that the processor's systems do not meet accessibility criteria must be able to raise a complaint to the processor (in line with the processor's complaints policy) or directly to the accrediting body.
- Any records produced by these systems are quality reviewed and assessed via regular and ad hoc audits. Findings of these audits are captured and translated into actions and lessons learnt.
- Management information must be produced through the interaction of researchers and processors via the various systems, reviewed and translated into actions.
- Support staff must be aware of their responsibility in delivering a quality service to researchers and that should be captured in their personal development and performance. Incidents related to poor or improper advice must be investigated and reviewed.
- Proactivity in the processors' approach is maintaining a set of quality of service(QoS) indicators to assess the service provided by systems (and support staff) and inform decisions and change initiatives improving the service offering.

C.4.4. Processor maintains clear and consistent records of all service users, including researchers' accreditation and training.

What to expect: A straightforward control requiring a register of all users, key accreditation, and training information. Information on this register must be consistently recorded and quality assured.

We expect data processors to:
- Maintain consistent records of all service users including researchers applying for accreditation, researchers undergoing training, accredited (fully and provisionally) researchers, supervisors/peer reviewers, as well as support staff accessing the data.
- These records are regularly audited, and quality reviewed, and findings of audits are translated into specific and time-bound actions.
- Instances where accreditation or access is suspended or withdrawn are clearly recorded, along with any incidents related to breach of accreditation conditions. These instances are reviewed regularly, and lessons learnt are produced and disseminated.
- Management information is produced, including a variety of quantitative and qualitative metrics, to inform decision on service provision. These metrics are used by staff to provide assurance that only accredited researchers and authorised staff interact with data, as well as evaluate the performance of the service.
- Observations and incidents relating to the accreditation of researchers (e.g., expiring, or expired accreditation or unauthorised access to data by staff) are recorded appropriately and actioned in line with relevant procedures, including escalation to the accrediting body where necessary. These procedures are reviewed and updated regularly.
- The proactivity of the data processor relies on maintaining and testing relevant procedures to proactively restrict access to data, and detect behavioural patterns related to non-compliance.

C.4.5. Each dataset is accompanied by a minimum set of documentation available to researchers and support staff. Processes to review data documentation are in place.

What to expect: We expect that the data processors will develop a minimum set of supporting information against each dataset curated in the environment. This information, would allow researchers to further understand data, including data restrictions and assumptions. As most environments do not enable access to publicly available resources (e.g., Quality and Methodology information) it is important that processors provide alternatives.

We expect that depending on the level of maturity processors will:
- Maintain and regularly review documentation against each dataset. The content and format of this documentation is up for the processor to decide but there is a requirement that the information is easy to understand, includes any

critical information that researchers need to make informed decisions and remains up to date.

- A process is in place to quality assure this information and feedback from researchers should be accounted for when producing and reviewing this material. The results of the quality audits are regularly reviewed and any actions to improve the quality of documentation have clear ownership and timelines.
- Processors have defined clear roles and responsibilities of staff responsible for producing, reviewing, and updating this information. Support staff advising on projects, curating data, and checking outputs should be aware of where to find it and be able to raise concerns or offer feedback.
- A mature processor would develop a consistent method in collating the required documentation and would ensure a consistent and accessible format and preferably held in a central repository.
- Management information pertaining to the quality and consistency of documentation is regularly reviewed and appropriately actioned.
- Proactive organisations have a clear process starting from the acquisition of data to translate information and metadata into this set of documentation. They would have also developed automated processes to extract key information from data (e.g., variable catalogues).

C.4.6. Processor provides access to policies and procedures on how researchers can interact with code repositories in the environment before researchers access the environment.

Optional

What to expect: Code repositories can provide a valuable resource to researchers and staff curating data, aiming to reuse code across different projects and datasets. However. there is need to administer these repositories, and to ensure that no information that would constitute a breach of the project accreditation conditions or a security breach is contained within them.

When applicable we expect that:
- Processors have clear and regularly updated policies and procedures in managing cross-project code repositories. This must include
  - well-defined roles and responsibilities for support staff administering these environments,
  - code standards and conditions of use,
  - incident reporting, and escalation routes, and
  - conditions for sharing and disseminating code from such repositories.
- Evidence is provided of regular and ad hoc audits on code repositories and compliance to all afore-mentioned procedures.
- Staff are aware of the risks involved in re-using code across project and have sufficient guidance, training, and support to review the contents of these repositories.
- Researchers are informed of their responsibilities and the terms and conditions of accessing and using, as well as sharing information from code repositories.

- Processors develop environment specific guidance to researchers developing code within the environment and using code repositories.
- Management information on code reusability and quality is produced and regularly reviewed to inform this element of service provision.
- Proactive organisations have in place adequate controls to prevent the dissemination of code repository content across projects without approval. They also maintain improved technical capacity to detect and review any instance of code extracted from code-repositories in the project environment. They seek to promote code re-use and offer specific training on the use of code-repositories in their environment.

Processor Accreditation obligations
How we interact with the accrediting body

This control area lists all obligations of the data processor to the accrediting body when accredited. It consists of controls relating to reporting requirements, and approval checkpoints by the accrediting body. Unlike previous categories there is no maturity assessment of a proactive approach for any of these controls. Maturity is assessed on the capability of the processor to produce timely and high-quality records, information and material supported by relevant management information. A summary of the information reported to the accrediting body is included in Appendix A.

C.5.1. Share relevant information on its performance.

What to expect: The processor must be able to provide key performance indicators(KPIs) as defined, in scope and format by the Accrediting Body. These can be supplemented by a narrative if the processors deems that it is necessary. The same performance information would need to be made publicly available (C.4.1) so this control does not introduce an additional burden to the processor. If the processor chooses, they can update the accrediting body with updated information as KPIs under C.4.1. are updated. The relevant KPIs are presented in Appendix A-Part D.

We anticipate that:
- Processors can produce KPIs within two weeks upon the request of the accrediting body to the quality standard expected.
- Processors maintain procedures to produce and update these KPIs as well as quality assure them.
- Management information is produced by the processor on its capability to produce these metrics to the standards of the accrediting body.
- The processor uses these indicators to inform decisions on its service and feedback is captured by support staff and users of the service on how efficiently these metrics capture the performance of the data processor.

C.5.2. Share relevant information on confirmed breaches and near misses by individuals and organisations.

What to expect: The processor must provide key information on incidents and near misses as defined in scope and format by the accrediting body. This information will relate to specific individuals (without disclosing the name of individuals) and organisations (researcher affiliations, project sponsors or third party when onward sharing data). The distinction between near misses and incidents is made clear in the relevant section of Appendix A (Appendix A-Part B).

We anticipate that:
- Processors can produce the requested information within two weeks upon the request of the accrediting body to the quality standard expected.
- Processors maintain procedures to produce, and quality assure this information.

- Management information is produced by the processor on its capability to produce this information to the standards of the accrediting body.
- The processor uses this information to inform decisions on its service and feedback is captured by support staff on how efficiently these metrics capture effect of security incidents to data capability.

C.5.3. The training course offered to researchers is recognised by the accrediting body and processors provide evidence that it is being regularly reviewed. Optional

What to expect: The processor must provide the access to or a copy of the content of the training course(s), any training plans, and syllabi of course(s) offered for the accreditation of researchers. The accrediting body will review these to ensure that key topics, as defined in scope and content by the accrediting body (Appendix A-Part F), are adequately addressed. Evidence of reviews of training material are also required to demonstrate the ongoing development of these courses. There is no expectation for the accrediting body to approve training material related to the running of the service (e.g., how researchers operate in the environment), but this can be provided as evidence of training under the people capability control (C.3.1.).

When training is offered, we expect that:
- Processors provide a copy of or access to the training material within two weeks upon the request of the accrediting body.
- Processors provide a copy of any supporting training material as well as evidence of reviews of training material to demonstrate the ongoing development of the training course.
- Any actions to improve the training material so as it meets the standards of the accrediting body have action owners and clear timelines.
- The processor captures feedback from researchers to inform the development of the training material, and this feedback is accounted for in supporting training material.
- Management information is produced by the processor on its capability to produce training evidence to the standards of the accrediting body, address improvement actions and on the quality of the training offered.

C.5.4. Share records of the data available in the processors' environment under the DEA.

What to expect: The processor must provide a list of data made available in the processor's environment under the Digital Economy Act 2017. This doesn't include data not accessed under the DEA. The process of reviewing this data catalogue for quality and consistency is a key requirement. Essential variables of the data catalogue are presented in Appendix A (Appendix A-Part C).

We anticipate that:
- Processors can produce consistently a DEA data register within two weeks upon the request of the accrediting body to the quality standard expected.

- Processors maintain procedures to produce, review, and quality assure this information.
- Management information is produced by the processor on its capability to produce a DEA data register to the standards of the accrediting body.

C.5.5. Share relevant information of all accredited researchers to the accrediting body.

What to expect: The processor must provide consistent records of all accredited researchers, as well as researchers undergoing the accreditation process, under the Digital Economy Act. The scope, format and content of this data is determined by the accrediting body (Appendix A-Part A)

We anticipate that:
- Processors can produce consistently a DEA accredited researcher register within two weeks upon the request of the accrediting body to the quality standard expected.
- Processors maintain procedures to produce, review, and quality assure this information.
- Management information is produced by the processor on its capability to produce a researcher register to the standards of the accrediting body.

C.5.6. Evidenced process to alert the accrediting body of any changes to a project that might impact the conditions of its accreditation.

What to expect: The processor must have a clear process and provide guidance to support staff to alert the accrediting body when any change to a project (change of scope, addition of data or data variables, extension of project, change of research team or affiliated organisations) can affect the conditions of accreditation.

There is a reasonable expectation that:
- The processor records and decides on change requests for project consistently and assesses the impact of changes against the conditions of project accreditation.
- A process to alert the accrediting body as and when required is in place and any alert to the accrediting body is recorded along with its decision.
- Guidance is provided to support staff managing change requests, and staff have access to information required to assess the extent and frequency of changes to a project. This guidance is regularly reviewed.
- When audits of the accrediting body disagree with the decisions made by the support staff approving change requests, a structured process is in place to record and investigate the incident and issue appropriate corrective action.
- Management information is produced and reviewed to manage the volume and scope of change requests against projects and researchers.

Assessment reviews

Under the capability framework there are three types of reviews of the control framework:
1. full accreditation reviews,
2. accreditation reviews, and
3. ad hoc audits

Full accreditation reviews take place every five years and consider all controls relevant to the functions of the data processor as well as any previously identified actions improvement actions. These reviews are conducted regardless of the maturity level of the data processor.

Regular accreditation reviews take place regularly and with a scope dependent on the maturity level of the accredited data processor. For processors assessed as maturing or mature, only controls assessed as maturing or capable are reviewed. For processors assessed as capable all controls will be reviewed. In addition to controls any previously identified improvement actions are also reviewed to ensure these are implemented. If improvement actions are not implemented at the time of the audit, the relevant control scores might be at risk of being downgraded. This is particularly important for controls that were maturing, as there is an expectation of enhancing capability by addressing the improvement actions.

Ad hoc audits are an instrument used sparingly to verify that controls remain capable after the previous accreditation. These are conducted only if there is reasonable doubt about specific controls (concerns). The scope of these audits are only controls assessed as capable. As a result of an ad hoc audit, additional improvement actions might be identified but the maturity assessment opinion will not change. Only if the ad hoc audit determines that the capability of the data processors against these controls is partial or minimal, the Research Accreditation Panel will be called to determine if the data processor can remain accredited.

At the end of the assessment, assessors will meet with data processors to discuss their findings, agree on the improvement actions, and share any concerns. This offers the opportunity to data processors to provide any further evidence and have clearer expectations at the next assessment review.

Assessments reviews will be accompanied by a report separated in two sections to cover both capability and security. The capability section will include:
1. a list the functions reviewed,
2. the type and scope of the review,
3. the maturity level opinion of the assessor against each control,
4. improvement actions and concerns against each control,
5. the weighted maturity level opinion of the assessor against each control area, and
6. the final recommendation of the assessor to the Research Accreditation Panel regarding the overall capability of the data processor.

The result of the assessment is in practice the independent opinion of the assessor, and it is ultimately up to the accrediting body to decide whether to:

- accept the assessor's opinion,
- issue amendments to the assessor's opinion without auditing the data processor,
- issue amendments to the assessor's opinion and a further review of the data processor's controls, or
- reject the assessor's opinion and a full review of the data processor's controls.

APPENDIX A

This appendix sets the reporting requirements to the accrediting body for specific controls in the data capability accreditation framework. These do not include any requirements as a result of the controls in the security accreditation framework. The reporting requirements differentiate depending on the scope of the accreditation (for the provision or preparation of the data) and the functions of the data processors.

Table of Contents

Part A - Information provided to the accrediting body regarding researchers

This includes information that related to researchers and their accreditation in response to **control C.5.5**. Each accredited researcher will be assigned a unique identification number provided by the accrediting body. This information will be retained by the accrediting body primarily for administrative purposes. The accrediting body might use this information to produce publicly available analysis in relation to its functions. Some of this information will be published in a public register of researchers.

It is important to note that when a researcher's accreditation is renewed, we expect processors to provide the latest dates. For instance, if a researcher was first assessed on 31 March 2018 after attending the Safe Researcher training course administered by the Secure Research Service but was then re-assessed on 10 July 2020 after attending the refresher course administered by the UK Data Archive, we would expect that his record would be updated to reflect the most current training.

The data processor has the right to object to the publication specific researcher information or the entire researcher record under specific exception conditions and only after approval by the Research Accreditation Panel. This information is still subject to access requests under the Freedom of Information Act (FOIA) and Freedom of Information (Scotland) Act (FOISA).

	Information	Format	Comments
Researcher essential information	Accredited researcher unique number		As provided by the UK Statistics Authority.
	Full name		This should include any middle names as recorded in official documents.
	Research affiliation(s)		If the researcher's access to the project is related to the work of a particular organisation this has to be registered. As researchers might work on multiple projects sponsored by different organisations all these affiliations will need to be recorded.
	Type of accreditation	Provisional/Full accreditation	
Training, assessment & accreditation	Training course		The name of the training course attended as part of the accreditation (e.g., Safe Researcher Training).
	Course provider		The processor who was responsible for delivering this course (e.g., Data Archive).
	Trained on	Day/Month/Year format (DD/MM/YYYY)	The date the researcher attended the training course.
	Assessed on		The date the researcher completed the assessment.
	Accredited on		The date the researcher was accredited.

Part B - Information provided to the accrediting body regarding incidents

This includes information that related to incidents and near misses in the processor's environment in response to **control C.5.2**. Processors might maintain their own incident investigation forms and provide them to the accrediting body. The information presented below should be collected in the processor's forms (in any format) to demonstrate an acceptable level of capability to manage incidents. This is not an assessment of the security of the environment and will not be assessed as such. Consequently, the data processor can decide on the applicable incident classifications, how to best assess the incident severity in their environment and describe how these recommendations are applied in their own environment.

This information will be retained by the accrediting body primarily for administrative purposes. The accrediting body might use this information to produce publicly available analysis in relation to its functions which might include aggregated processor information.

	Information	Recommended format	Comments
Basic incident information	Incident summary	Up to 500 words	
	Data Processor		
	Incident severity	Major incident Minor incident Near miss	*See incident severity section*
	Incident classification		How would you classify this incident? *See incident classification section*
	Detected via		How was the incident detected? *See event detection method section*
	Incident date	Date/Month/Year Hour/Minute format (DD/MM/YYYY hh:mm)	
	Incident reported to data processor on		
	Incident reported to accrediting body on		
	Incident resolved on		
	Summary of actions by the processor to respond to the incident	Up to 500 words	
Data incident investigation	Was data confidentiality compromised?	Yes/No	
	Did any data leave the processors environment?	Yes/No	This includes data leaving in the environment in any format (electronic of physical).

	Does this require access to data to be suspended or revoked?	Suspended/Revoked/Neither	
Researcher incident investigation	Accredited Researcher(s)	Accredited researcher(s) unique number	If applicable
	Number of previous incidents in relation to the researchers		
	Accredited project	Accredited project unique number	
	Number of incidents in relation to this project		
	Does this require researchers to be retrained?	Yes/No/N/A	
	Does this incident affect the accreditation of researchers?	Yes/No	
	Justification of the previous response		
Internal incident investigation	Is this a staff related incident?	Yes/No	If applicable
	Number of incidents in relation to the members of staff involved		
	Is this incident related to a particular policy	Yes/No	
	Name(s) of related policies		
	Number of incidents in relation to the particular policy		
	Is this incident related to a particular activity/procedure	Yes/No	
	Name(s) of related activities/procedures		
	Number of incidents in relation to the particular procedure		
Data processor incident investigation	Does this incident affect the accreditation of the processing environment?	Yes/No	
	Justification of the previous response		
	Does this incident require a review of the Data Protection Impact Assessment?	Yes/No	
	Justification of the previous response		

Incident severity

Incident severity is defined differently in various environments. In terms of reporting, we expect that any incident that has happened in the processor's environment is classified as either major or minor. Examples of incidents assessed as major are:

  i.   data confidentiality has been breached (in line with the data confidentiality policy),
 ii.   data have been exported outside the processor's environment either in an electronic (e.g., a researcher or support staff export non-disclosure controlled data) or a physical format (e.g., a researcher makes a copy of information from the processor's environment on a notepad),
iii.   identifiable or illegally acquired data have accessed in the processor's environment and statistical outputs involving this data have been produced (regardless of the granularity of outputs),
 iv.   research methods indicate that the researcher attempts to re-identify data within the environment,
  v.   statistical outputs outside the scope of the project have been published, and
 vi.   the security controls of the environment have been compromised.

Examples of incidents assessed as minor are:

  i.   identifiable data have accessed in the processor's environment but not used in the production of statistical outputs,
 ii.   statistical analysis is outside the project scope, but no outputs have been disseminated,
iii.   researchers or staff don't comply with data capability policy requirements and processing environment procedures,

A near miss is an event relates to an incident that has been avoided by chance and not via an established control. For instance, support staff clearing an output realise that a researcher has been given access to variables that they shouldn't not have access to. This is not a regular check for statistical disclosure control but at data ingestion.

If a particular researcher, member of support staff, policy or procedures are systematically involved in minor incidents or near misses this ought to trigger a major incident. This is why we expect a mature processor to record and report more granular incident information.

Incident classifications

Different processors will classify incidents in different ways and in many cases an incident might not be attributed to a single cause. Recording and reporting incident causes can offer valuable insights for processors and enable the accrediting body to realistically assess the evidence provided against the accreditation criteria and issue meaningful improvement actions. Contrary to the many assumptions, a mature processing environment is not necessarily an environment where no incidents or near misses occur. It is an environment where incidents are identified and addressed quickly and efficiently. A processor with limited capacity to identify minor incidents is inevitably a processor that will be ill-equipped to manage a serious incident.

We have listed below some common classification of incidents and near misses which might be found in a processor's environment:

  • unauthorised access to data,
  • use of data outside project scope,

- unauthorised dissemination of data,
- dissemination of de-identified or identifiable data,
- improper request to access researcher data,
- non-compliance to policy,
- attempt to re-identify data,
- compromise of access credentials,
- access from an unauthorised access location/point,

These lists are not inclusive but a mere suggestion of types of incidents.

Event detection method
Technology, with the use of complex machine learning algorithms and systems capable of producing a more detailed audit trail, offer us several ways to detect incidents and near misses before it is too late. Traditional methods as direct reports from users and researchers and sample audits are also effective. The capability of the processor and ultimately the security of the environment depends on the efficacy of these methods. Consequently, it is important for processing environment to know how indents come to their attention. Our suggestions for different generic detection methods are presented below:
- event reported internally (e.g., researcher, support staff)
- event reported externally (e.g., member of the public)
- event detected through a sample audit
- event detected via technology-driven monitoring mechanisms
- event detected as a result of another incident investigation

Part C - Information provided to the accrediting body regarding data accessed via the Digital Economy Act 2017 legal gateway

Each processing environment would define its own metadata model when it comes to cataloguing data. This annex suggests some essential information that must be covered by any data catalogue as part of control C.5.4.

This information will be retained by the accrediting body primarily for administrative purposes. The accrediting body might use this information to produce publicly available analysis in relation to its functions which might include aggregated processor information.

The data processor has the right to object to the publication specific data information or the entire data record under specific exception conditions and only after approval by the Research Accreditation Panel. This information is still subject to access requests under the Freedom of Information Act (FOIA) and Freedom of Information (Scotland) Act (FOISA).

	Information	Description	Format	Comments
Essential information	Data name	A unique name provided to identify the data		
	Data description	A short description to what this dataset involves (can include external links)	Up to 200 words	
	Data classification	The type of data (e.g., household survey data, administrative data, open data)		
	Data keywords	A set of keywords that relate to the data		
	Data supplier	Who is the supplier of the data? For personal data please provide the name of the data controller.		
Data details	Time coverage - start	The first month the data covers	At least month and year (MM/YYYY)	
	Time coverage – end	The latest month the data covers		If there is an ongoing supply of data, data suppliers must specify that this is ongoing.
	Data frequency	In case the data can be separated in data intervals please provide the data frequency (e.g., a quarterly survey will have a quarterly frequency)		For ongoing data supplies, please specify how frequently will data be updated in the environment. For instance, a quarterly survey might provide data twice a year.

	Geography	The levels of geography included in the data		This includes all levels of geography excluding levels accessible via lookups. For instance, if the data contains a postcode variable there is no need to include higher levels as region which can be obtained by using a lookup.

Part D - Information provided to the accrediting body regarding performance

This includes information related to the performance of the different processors. Given that not all processors are expected to deliver all functions, it is up to processor to identify which performance indicators relate to their own functions. This is in response to control C.5.1.

This information will be retained by the accrediting body primarily for administrative purposes. The accrediting body might use this information to produce publicly available analysis in relation to its functions. Some of this information will be published.

Many of the key performance indicators listed below depend on multiple parties within the research community, as researchers and data providers. When producing these metrics there is a caveat that delays can be caused in the processors' environment outside their control. It is also important to present and examine the trends of these metrics over time and provide any supporting narrative.

Key performance indicator	Time to make data available
Functions	Preparation of data Provision of data
Purpose	Provide researchers a realistic timeline for making data available in the processors' environment. This will enable researchers to make informed decisions to plan their research project and data suppliers on how they can enable timely access to their data.
What is measured	For processors that <u>are accredited to acquire data</u> we expect the following metrics: The average, median, minimum, and maximum time to acquire data from the time that a request to acquire is agreed to the point that the data, and any supporting documentation, is received in the processor's environment. This must be broken down by data supplier. For mature organisation this should be broken down by data theme (data keywords). For processors that <u>prepare data</u> we expect the following metrics: The average, median, minimum, and maximum time to prepare data from the time that the data is received in the processor's environment to the point that the data, and any supporting documentation is signed off as prepared. This must be broken down by data supplier. For mature organisation this should be broken down by data theme (data keywords). For processors that <u>provision data</u> we expect the following metrics: The average, median, minimum, and maximum time to provision data from the time that the

	project is accredited in the processor's environment to the point that the data, and any supporting documentation is made available in the project area. This must be broken down by data supplier and project sponsor. For mature organisation this should be broken down by data theme (data keywords) and project theme (project keywords).
Minimum frequency	Biannually

Key performance indicator	Data preparation efficiency
Functions	Preparation of data
Purpose	Provide researchers a realistic timeline for making suitable data available in the processors' environment. This will enable researchers to make informed decisions to plan their research project and data suppliers on how they can enable timely access to their data. As processors prepare data it is likely that they are asked to revisit a previously prepared dataset. The volume and frequency of these requests might indicate challenges regarding a specific dataset or the processor's processes. This is an indication on whether data processor's environment is capable of preparing a higher volume of data but not at a level of quality that would be useful to researchers, leading to repeated requests to review the same data.
What is measured	For processors that <u>prepare data</u> we expect two metrics: • the number of requests to prepare data, and • the number and the time between repeat requests (for the same dataset to be further prepared, including the provision of additional data/variables or revision of the documentation)
Minimum frequency	Biannually

Key performance indicator	Time to accredit researchers
Functions	Provision of data
Purpose	Provide researchers a realistic timeline for the researcher journey in processor's environment. This will enable researchers to make informed decisions to plan their research project and which processor to use for their accreditation. The accreditation process includes several steps that depend on multiple parties, including the accreditation body. The scope of these metrics is to limit what is measured to the functions of the processor.

What is measured	For all processors that accredit researchers we expect the following metrics: The average, median, minimum, and maximum time to do the accreditation checks from the time that an accreditation application is received to the point that it is sent to the accredited body for approval. For processors that <u>are accredited to provide training</u> we expect that the following metric is also included: The average, median, minimum, and maximum time to provide training, from the time that the request for training has been received to the point that the training has been delivered. For processors that are <u>not accredited to provide training</u> but depend on another environment we expect the following metric: The average, median, minimum, and maximum time to provide training, from the time that the request for training has been received to the point that the training has been delivered. This needs to be broken down by the training provider.
Minimum frequency	Quarterly

Key performance indicator	Processing project applications
Functions	Provision of data
Purpose	Provide researchers a realistic timeline for how long it takes to produce the project application. This will enable researchers to make informed decisions on the level of support offered by the processor's environment. The project accreditation process includes several steps that depend on multiple parties, including researchers and decision makers. These metrics are subject to delays caused by parties other than the data processor.
What is measured	For processors that accredit projects we expect the following metrics are reported: • The average, median, minimum, and maximum time from the time that researchers start their project application to the point that it is sent to the accredited body for approval. • The number of projects requesting to be accredited These must be broken down for each project sponsor. For mature organisations these should be broken down by project theme (project keywords).
Minimum frequency	Quarterly

Key performance indicator	Releasing data from the environment
Functions	Provision of data
Purpose	Provide researchers a realistic timeline for how long it takes to apply statistical disclosure controls and release them from the processor's environment. This will enable researchers to make informed decisions on the level of support offered by the processor's environment, to meet their publication schedule and decide which environment to use. The project accreditation process includes several steps that depend on multiple parties, as researchers and decision makers. These metrics are subject to delays caused by parties other than the data processor.
What is measured	For processors that accredit projects we expect the following metrics: • The average, median, minimum, and maximum time from the time that researchers submit their analysis for clearance to the point that it is released from the environment. • The volume of output requests should be also measured. These must be broken down for each project sponsor. For mature organisations these should be broken down by project theme (project keywords).
Minimum frequency	Quarterly

Key performance indicator	Number of service complaints
Functions	Preparation of data Provision of data
Purpose	In case users of the service are not satisfied by the quality of service offered in the processor's environment, they must be able to file a complaint. This can be directed to the processing environment or directly to the accrediting body. Measuring the volume of complaints is a crude indicator of the service provided.
What is measured	Volume of complaints and time to resolve a complaint, from the time the complaint was received to the point that the complainant agreed that it can be closed, by the data processor.
Minimum frequency	Annually

Part E - Information provided to the accrediting body regarding research projects

This includes information on projects, approvals, researchers, data, and published outputs. This information is required as part for controls C.1.1. and C.1.5. The requirement to report this information apply only to processors accredited for the provision of data. Each project will be assigned to a unique project identification number provided by the accrediting body.

This information will be retained by the accrediting body primarily for administrative purposes. The accrediting body might use this information to produce publicly available analysis in relation to its functions. Some of this information will be published in a public register of projects.

The data processor has the right to object to the publication specific project information or the entire project record under specific exception conditions and only after approval by the Research Accreditation Panel. This information is still subject to access requests under the Freedom of Information Act (FOIA) and Freedom of Information (Scotland) Act (FOISA).

	Information	Description	Format	Comments
Project information	Project title	The official project title as approved	Up to 100 characters	
	Project abstract	A short paragraph summarising the purpose of the project	Up to 250 words	
	Expected public benefits	A short paragraph summarising the expected public benefits		
	Project keywords	A set of keywords describing the project		For instance: employment, economy, crime, rehabilitation, environment
	Project start date	The date this project started in the data processor's environment	Day/Month/Year format (DD/MM/YYYY)	A project is considered to start in a processors environment when researchers have access to the environment and all data as approved in the project application. It is not the date of project approval.

	Project end date	The expected end date of the project		
	Research environment	The name of the accredited processor's environment where research will take place.		
	Research sponsor	The name(s) of the organisations sponsoring this research.		A project sponsor is the organisation actively supporting this research. This support can be financial (e.g., funding), reputational (e.g., affiliation with or endorsement of the project)
Approvals	Project approval on	The date this project was approved by UK Statistics Authority	Day/Month/Year format (DD/MM/YYYY)	
	Ethical approval on	The date ethical consideration/approval was given to this project		
	Ethical approval by	Who provided ethical approval for the project	The name of the entity that provided ethical approval	
	Ethical restrictions	Any restrictions identified as part of the ethical approval		
People	Research Lead (accredited researcher number)	The accredited researcher unique number for the Research Lead	A standard formatted number	The accrediting body will ensure the production of unique accredited researcher numbers upon researcher accreditation.
	Researchers (accredited researcher numbers)	The accredited researcher unique number for the Research Lead		
	People restrictions	Any restrictions on the people involved in this project identified as part of the project accreditation		

Data	Data used	A list of all datasets used in this project, including any data brought in by researchers		In case of exceptions in the publication of data used apply, these need to be communicated to the accrediting body.
	Data restrictions	Any restrictions on the data available to the project identified as part of the project accreditation		
Dissemination	Published research links	Links to published research (e.g., papers, articles, blogs) after project closure.	All links need to be publicly accessible and not behind a paywall.	
	Dissemination restrictions	Any restrictions on the dissemination of research outputs identified as part of the project accreditation		

Part F - Essential requirements for accreditation training course offerings

The accrediting body is responsible for approving any training course designed for external users accessing the processor's environment. This applies only to data processors that provision data and provide training to researchers and other users of the service. This relates to control C.5.3.

The main objective of the training is to equip users of the service with enough skills and knowledge on how to meet the conditions of their accreditation throughout the researcher journey. The assessment of any training offerings will be based on this objective and will be examined on a case-by-case basis. The training itself is not focused on the technical capacity of a particular environment but on the wider capacity of researchers to understand the purpose and significance of the accreditation conditions for researchers and projects.

From a data capability perspective any training course should expand on the following content:
  i.   Setting the right scope for your project and deviating from the original project scope
  ii.   The importance of ethical, policy and legal considerations in research
  iii. Assess and mitigate the risk of re-identification during analysis and dissemination of outputs
  iv.   Understand the risks and benefits of using multiple datasets, including bringing your own data and using open data, introducing new methods and algorithms, sharing code in a secure environment (this extends to security, ethics, and methodology).
  v.   Highlight the importance of safe statistical outputs in publicly available statistical research
  vi.   Explain how researchers' behaviours and actions can translate into security and ethical risks
  vii. How the researchers and other users can interact with the processor's environment and the role of the accrediting body, including contact information to the accrediting body to submit a complaint.

ANNEX A

This annex serves as a short glossary for terms found in the guidance which warrant additional explanation.

Software module
The term software module refers to software that is licensed, with or without a cost, as an addition to a larger software product.

Data sensitivity
Most organisations have developed or adopted a wide variety of frameworks used to assess the risk of data management activities. These approaches are tailored to the data strategy and risk appetite of each organisation and are appropriate for the functions they undertake.

One part of these approaches is concerned solely on the data and its inherent risk. Structured approaches or models assess key characteristics of data to determine the level of risk. The level of risk of a dataset is referred to as data sensitivity.

Information governance approaches use risk management practices to evaluate the level of risk. The assessment for each dataset is captured in information risk assessments.

Data extract
A data extract is dataset derived from another dataset (parent dataset) by reducing the information contained in the latter. This reduction can be seen as:

- the data extract does not contain all variables found in the parent dataset
- the data extract has reduced the detail in one or more variables (e.g., higher geography level, categorical or ordinal variables produced)
- the data extracts do not contain all records found in the parent dataset (e.g., smaller time frame).

Data extracts can be also referred to as data slices, data views, or data partitions.

Onward sharing
With the term onward sharing we refer to any dissemination of data, excluding statistical disclosure controlled statistical outputs, outside the data processor's environment. For example, this can entail a processor de-identifying a dataset or linking several datasets and then sharing them to another processor for the provision of data to researchers.

For the purpose of the accreditation, environments accredited for both the preparation and provision of data are treated as one processor, despite the requirement for these functions to be strictly separated. As a result, they do not need to report and review data transfers from the processing area to the researchers' area.

Physical and virtual data access

There are two concepts of data access in any environment; physical and virtual. The most common method is physical access when a data supplier sends a copy of a dataset, as a file or set of files, to the data processors environment. This is commonly known as data ingestion or data egress.

Virtual access refers to an access arrangement allowing a data supplier to provide to the data processor access to a dataset in the data supplier's own environment. The data process cannot tamper with the dataset but can allow access to it to specific users. This method is known as data virtualisation, as the data is only virtually in the data processor's environment.

ANNEX B

This annex outlines the different scenarios for the data capability accreditation reviews. This includes regular accreditation reviews depending on the maturity assessment opinion and the 5-years' full accreditation (filled in a light blue colour). The frequency of accreditation exercises in these five years is shown in the last column which is colour-coded appropriately. Processors can use this table by changing the year of their first accreditation and plan accordingly.

2022	2023	2024	2025	2026	2027	Number of accreditation exercises in 5 years
Mature			Mature		Mature	
					Good	
					Capable	
			Good		Mature	2
					Good	
					Capable	
			Capable	Mature/Good/Capable	Mature	
					Good	3
					Capable	
Good		Mature			Mature	
					Good	2
					Capable	
		Good		Mature/Good/Capable	Mature	
					Good	
					Capable	
		Capable	Mature/Good		Mature	3
					Good	
					Capable	
			Capable	Mature/Good/Capable	Mature	
					Good	4
					Capable	

2022	2023	2024	2025	2026	2027	Number of accreditation exercises in 5 years
Capable	Mature			Mature/Good/Capable	Mature	3
					Good	
					Capable	
	Good		Mature/Good		Mature	
					Good	
					Capable	
			Capable	Mature/Good/Capable	Mature	4
					Good	
					Capable	
	Capable	Mature			Mature	3
					Good	
					Capable	
		Good		Mature/Good/Capable	Mature	4
					Good	
					Capable	
		Capable	Mature/Good		Mature	
					Good	
					Capable	
			Capable	Mature/Good/Capable	Mature	5
					Good	
					Capable	