DATA FUTURES WHITE PAPER

HESA

Contact details
Jisc
95 Promenade
Cheltenham
GL50 1HZ
E data.protection@hesa.ac.uk
T +44 (0)1242 388 513 [option 5]
W www.hesa.ac.uk

CONTENTS

INTRODUCTION

Data Futures is a sector-wide transformation programme which has reformed the collection of higher education (HE) data by utilising recent technological innovation to advance the approach to data collection, assurance, and dissemination of HE data across the sector.

The programme achieved this by delivering a new data collection platform, the HESA Data Platform (HDP), which enabled the provision of data to statutory customers and funders for the Student Record. In addition, the HDP delivered operational efficiencies for HE providers to submit and quality assure their data.

The programme was initially led by the Higher Education Statistics Agency (HESA) in collaboration with its technical delivery partner Jisc and was shaped by engagement from HE providers and Statutory Customers and Funders.

The HDP delivers a wide range of benefits for statutory customers and funders, HE providers, Jisc, and the wider HE sector.

This paper provides an overview of the Data Futures programme and its approach to Data Protection and Information Security by design.

GLOSSARY

Glossary Term	Definition
By Design and Default	Organisations are encouraged to implement technical and organisational measures, at the earliest stages of the design of the processing operations, in such a way that safeguards privacy and data protection principles right from the start ('data protection by design'). By default, organisations should ensure that personal data is processed with the highest privacy protection (for example only the data necessary should be processed, short storage period, limited accessibility) so that by default personal data isn't made accessible to an indefinite number of persons ('data protection by default').
Check-in, Check-out	The implementation of a "check in – check out" system ensured the proper Privacy and Security considerations were identified, tracked and signed off in a time frame that is appropriate to the backlog item.
Gateway Document	Single point of truth document, maintained by the Information Security and Data Protection SME to track the requirements and clarify if they have been met.
Lawfulness, Fairness and Transparency	Any processing of personal data should be lawful and fair. It should be transparent to individuals that personal data concerning them are collected, used, consulted, or otherwise processed and to what extent the personal data are or will be processed. The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used.
Purpose Limitation	Personal data should only be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes. In particular, the specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data. However, further processing for archiving purposes in the public interest, scientific, or historical research purposes or statistical purposes (in accordance with Article

Glossary Term	Definition
	89(1) GDPR) is not considered to be incompatible with the initial purposes.
Minimisation	Processing of personal data must be adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed. Personal data should be processed only if the purpose of the processing could not reasonably be fulfilled by other means. This requires, in particular, ensuring that the period for which the personal data are stored is limited to a strict minimum.
Storage Limitation	Personal data should only be kept in a form which permits identification of data subjects for as long as is necessary for the purposes for which the personal data are processed. In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review.
Accountability	Accountability is a common principle for organisations across many disciplines; the principle embodies that organisations live up to expectations for instance in the delivery of their products and their behaviour towards those they interact with. Data Protection legislation integrates accountability as a principle which requires that organisations put in place appropriate technical and organisational measures and be able to demonstrate what they did and its effectiveness when requested.
Providers	The umbrella terms higher education provider (HE provider) and provider are used to describe the organisations from which HESA collects data.
Statutory Customers	Data Collection and Statistics directorate (DCS) shares information with public authorities who require it to carry out their statutory and/or public functions.
Epics	In agile development, an epic represents a series of user stories that share a broader strategic objective.
Backlog item	A product backlog is a prioritised list of work for the development team that is derived from the roadmap and its requirements.
User Stories	A user story is a tool in Agile software development used to capture a description of a software feature from a user's perspective. The user story describes the type of user, what they want and why. A user story helps to

Glossary Term	Definition
	create a simplified description of a requirement.
Data Subject	The identified or identifiable living individual to whom personal data relates.
Controller	The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
Processor	The natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. Processors act on behalf of the relevant controller and under their authority.
Product Owners	A product owner is a role on a Scrum team that is responsible for the project's outcome. The product owner seeks to maximise a product's value by managing and optimising the product backlog.

mission and principles that guide the programme

HESA was set up by agreement between HE funding councils, HE providers, and the relevant government departments. The Further & Higher Education Act 1992 and the White Paper 'Higher Education: a new framework' had identified a need for a co-ordinated approach to higher education statistics and information. In 2018, HESA was confirmed as the DDB for higher education in England.

HESA's mission was to support the advancement of UK higher education by collecting, analysing, and disseminating accurate and comprehensive statistical information in response to the needs of all those with an interest in its characteristics and a stake in its future.

Jisc and HESA merger

In October 2022, the two UK sector agencies Jisc and HESA merged. As part of the union, the Department for Education transferred HESA's status as England's Designated Data Body (DDB) to Jisc.

Existing HESA staff joined Jisc as a new Data Collection and Statistics (DCS) directorate, continuing the delivery of the Data Futures programme. Jisc, which is UK higher education's main technology organisation, has subsequently become solely responsible for the successful delivery of the HDP into a production environment.

The DCS directorate within Jisc will continue to collect and analyse data and publish statistics under the HESA brand. The DDB continues to be driven by the same HESA mission and is guided by Jisc's strategic vision for 2023 – 2025 to leverage the collective power of the sectors to maximise impact. It will achieve this by committing to the delivery of the right solutions, empowering communities, and being a force for good.

Jisc's DCS deploys its expertise in support of this vision in the use and application of advanced statistical and open data techniques to transform and present HE data by developing innovative, low-cost techniques to improve the quality and efficiency of data collection. Jisc will also work as part of an open infrastructure to ensure that as much data as possible is open and accessible to all.

Underpinning Jisc's core purpose to collect, assure, analyse, and disseminate data and information on all aspects of UK higher education are information security and data protection principles that have been designed, embedded, and upheld across all DCS processing.

Programme approach

HESA and Jisc established positive and effective ways of working from the outset of the Programme. The approach to delivery involved mixed development teams working in line with agreed agile ways of working. This approach allowed for the most effective use of skills and resources, with rapid testing of systems and components, and re-ordering of priorities at short notice if required.

As part of the commencement of the programme, a requirement gathering exercise was completed. This included breaking down the fundamental elements of the end-to-end delivery, splitting this into high level and detailed requirements. This activity was completed by the relevant Subject Matter Experts (SMEs), including Information Security and Data Protection personnel, and signed off by the business owners. This list of requirements was reviewed by the Programme's Product Owners to assist with splitting into Epics and User Stories that would ultimately make up the backlogs for the development teams. The Epics and associated User Stories were then prioritised by the Product Owners to create the Programme Roadmap ('Roadmap').

The Roadmap set out the key deliverables aligned to Alpha, Beta and Transition milestones. The Roadmap was broken down into increments, with more detailed planning to take place at the start of each increment, including building in any learning from the previous stage.

In line with agile practices, the backlog could be re-ordered within increments to prioritise delivery effectively, with changes being reported and agreed at the Data Futures Delivery Group (DFDG) to ensure alignment across the programme. Re-ordering across increments was in agreement with the Data Futures Programme Board (see more: Engagement and Programme Governance).

Information Security and Data Protection considerations were incorporated from the outset, with ongoing involvement from DCS specialists in this area, and the ongoing development of the Data Protection Impact Assessment (DPIA). This approach ensured that Data Futures is delivered to a standard compliant with applicable Data Protection legislation and Information Security standards.

To this end, the Programme ensured the allocation of dedicated Information Security and Data Protection Compliance Officers (SMEs) from its inception to the Production release. These individuals will continue to support and offer their expertise to the Product Owners following the transition to business as usual.

Timescales for feature development were driven by the optimum timeframes for Alpha and Beta to ensure providers were able to participate without this contending with their existing Student or Student Alternative data returns. Alpha took place between April and July 2021, with Beta running

between February and November 2022. Roadmaps were designed to support completion of development in time to allow penetration testing and any remediation required prior to the start of each testing phase.

All development teams used JIRA to manage development tasks, present progress during sprint reviews and provide reports for key stakeholders. Progress on JIRA user stores were fed directly into the roadmap to show what has been achieved against it.

EMBEDDING DATA PROTECTION AND INFORMATION SECURITY

Information Security and Data Protection requirements were at the heart of the design, decision-making, and build of the HDP. The programme achieved this by introducing several Information Security and Data Protection enhancing measures to ensure that all elements of the delivery were considered from an Information Security and Data Privacy by Design and Default perspective.

THE DATA FUTURES JOURNEY

Requirements Gathering

The following is a summary of the Data Privacy and Information Security requirements that were identified before the commencement of the Data Futures Programme:

Summary of common requirements

1. The product must support HESA's continued compliance with relevant security related certifications and legal obligations including:

    a. Data Protection Act 2018 (DPA 18)
    b. UK General Data Protection Regulation (GDPR)
    c. ISO 27001

2. The system design should incorporate relevant vendor best practice.

3. The design must reflect a defence in depth approach which utilises relevant access controls including perimeter firewall, logical network segmentation, Network Security Groups (NSGs), routing control, IP and port filtering, Role Based Access Control (RBAC), IaaS VM firewall and Anti Malware defences.

4. The documentation must support security incident investigation and resolution which may be required remotely outside of normal business hours. The documentation must therefore be accessible in a format and location which may differ from standard operational documentation.

5. Data protection rights should be enabled, where valid requests are made.

6. Notices of consent and fair processing must be updated to maintain data protection compliance. See https://www.hesa.ac.uk/about/regulation/data-protection/guidance.

7. Where security requirements are not complied with, for example due to system limitations, performance, usability, or cost trade-offs then a detailed justification and possible mitigations must be included in system documentation and agreed as part of the governance process. This

is an addition to the requirement to justify non-compliance to ‘must have’ requirements. This reflects the increased level of justification required for security related issues.

Confidentiality and integrity

Multi factor authentication (MFA) must be considered during design, based on risk assessment.

Use cases where MFA should be the default choice are:

a. System administration access
b. First time user access
c. High risk or privileged access such as changing passwords and identity details
d. Virus and malware checking for uploaded files must be considered during design based on risk assessment
e. All data at rest must be encrypted
f. All data in transit must be encrypted
g. All inbound network traffic must traverse a Web Application Firewall i.e. offers protection against attacks using cross site scripting, SQL injection, HTP protocol anomalies, denial of service
h. Keys and passwords must not be stored in plain text e.g. in configuration files or source files

Non-repudiation and accountability

All actions by a user or system which create, retrieve, modify, or delete data must be logged.

Logged information must include:

a. Event time e.g. UTC timestamp
b. Identity of the system or user e.g. user ID, IP address
c. Target of the request/action e.g. system, service, type of store
d. Description of data accessed, or action attempted /performed
e. Success or failure or action

Event times must be co-ordinated throughout the system so that events across the system services and components can be reconstructed into a post event timeline. The reconstruction of the timeline should not require developer resources.

ALPHA

Alpha was the first phase of the Data Futures programme. The purpose of this phase was to test the new HDP concepts that had previously been unavailable as part of the DCS’s historic collection process, such as the use of 'tolerances' in Issue Management.

The Alpha phase included a small cohort of 14 Providers and four Statutory Customers. This phase was governed by a set of ‘Programme Gateways’ to control any Data Protection and Information Security risks arising from the initial development of the HDP processing.

Introduction of Programme Gateways

The Gateway Framework introduced a set of compliance sign-offs aligned with the go-live dates for each increment of the Alpha and Beta phases. Sign-off was required by HESA's Data Protection Officer, Chief Technology Officer, and General Counsel before the programme could progress through the gateways and deploy any new developments to the test environments.

This approach was supported by a single point of truth Gateway Document, maintained by the Information Security and Data Protection subject matter experts deployed to the Programme. This document acted as point of risk escalation and yielded bespoke requirements that the Programme was obliged to fulfil before it could progress to the next increment.

As the programme progressed, risks were identified in a timely manner and added as a requirement to the Gateway Framework. If there were any perceived high risks, the Gateway Framework ensured that these were mitigated before the Programme could progress. At the end of every release phase sign off was provided by the Programme to confirm that requirements had been met.

BETA

As the Information Security footprint and Data Protection risks for Beta were significantly increased over those for Alpha, owing largely to the potential processing of personal data and the estimated inclusion of 100 Participant Providers, it was crucial that Data Protection and Information Security concerns were identified and addressed in a timely fashion.

To this end, the Programme ensured that controls were in place to ensure Providers could safely use real personal data in the HDP during the Beta phase. The environment was set up to ensure the secure processing of personal data, as if it were live and being used in the business-as-usual Student Record collection.

As part of the Programme's continual improvement process, it introduced a new governance framework for maturing the assessment of Data Protection and Information Security risk in line with the programme's Agile Methodology.

To this end, the subject matter experts worked with key stakeholders to define a Check-in, Check-out process to review Data Protection and Information Security considerations for each individual Epic. This matured the Programme's approach to embedding Security and Privacy by Design and Default into the initial stages of the software development processes.

Check-in, Check-out Approach

The implementation of a Check-in, Check-out approach further ensured that Data Protection and Information Security considerations were identified, documented, tracked, and signed off in a time frame that was appropriate to the Epic.

This approach continued to guarantee that compliance considerations were identified before development activities started, were tracked as those activities progressed, and were signed off before the Epic functionality was moved into a deployment release.

To help identify relevant requirements for each backlog item, a standard checklist of questions was formulated with the intention that these were applied to each backlog item at the outset and revisited periodically throughout the item's lifecycle to the point of sign-off and delivery.

Refer to Annex 1 to see the full checklist used to assess each Epic.

Check-in

Backlog items that had Data Protection and Information Security requirements were identified at an early stage via a workshop between the SMEs and technical and management stakeholders of the development teams.

The Gateway Framework was used to document, track and update acceptance criteria for each backlog item throughout its lifecycle. The Gateway Document was informed by the responses given to the Data Protection and Information Security Requirements questions as part of each Epic checklist.

At points of significant change in the backlog item, the requirements were reviewed by the Product Owner or relevant Project Management Office team member(s) together with the SMEs and the Gateway Document reassessed to ensure accuracy and applicability.

Any material changes to the requirements were authorised by the Data Protection Officer, Chief Technology Officer, and General Counsel, as appropriate.

Check-out

Epics were subsequently signed off by the SMEs upon agreement that all requirements within the compliance requirements were met. Sign off from these individuals needed to be obtained before the Epic could reach the end of its development lifecycle. The SMEs satisfied themselves that requirements had been met through a combination of demos, review of documentation and detailed discussion with architects, developers and other stakeholders.

The Check-in, Check-out approach was implemented since the beginning of Beta and was followed for all development activity leading up to the production release. The approach was influential in supporting the programme's approach to undertaking a programme wide Data Protection Impact Assessment through its continued endorsement of Information Security and Privacy by Design and Default.

DATA PROTECTION IMPACT ASSESSMENT (DPIA)

UK data protection legislation requires Controllers to put in place:

- appropriate technical and organisation measures to implement data protection principles effectively; and
- measures to safeguard individual rights

To achieve this, the Programme 'baked in' data protection into its processing activities and business practices, from the design stage and throughout the processing lifecycle of personal data.

A DPIA is a process designed to help organisations systematically analyse, identify, and minimise the data protection risks of a project or plan. It is a key part of the DCS's accountability obligations under the UK GDPR and has assisted the programme in assessing and demonstrating how it complies with all its data protection requirements.

The DCS is legally required to undertake a DPIA before beginning any type of processing that is "likely to result in a high risk". This means that although the actual level of risk may not yet have been assessed, factors that point to the potential for a widespread or serious impact on individuals need to be identified and mitigated before the proposed processing begins.

Following the UK GDPR and ICO requirements, Jisc is therefore required to perform a DPIA if it undertakes the following:

- Processing of special category or criminal offence data on a large scale
- Use of innovative technology
- Matching data or combining data sets from different sources
- Processing data that might endanger the individual's physical health or safety in the event of a security breach

Given the criteria outlined above, the DCS Data Protection Officer considered that all development aspects would ultimately be in scope and therefore the Check-in, Check-out approach enabled the SMEs to review and assess the impact of each Epic. The output of each checklist was used to inform the DPIA.

HESA approached the completion of the DPIA as an ongoing process which was embedded in the development of the Programme and therefore subject to regular review. Consequently, given that this was an agile programme with multiple releases, HESA's DPIA continued to evolve alongside the progression of the programme and covered the following elements:

- processing overview
- consultation process
- whether the processing achieves the intended purposes and whether there is another way to achieve the same outcome
- how purpose limitation and prevented function creep is ensured
- steps taken to ensure data minimisation, and that only the data necessary for the project is processed
- steps taken to ensure data quality, and that information is kept up to date and checked for accuracy and completeness
- how individuals' data privacy rights are upheld
- the contractual arrangements in place to facilitate ingestion or onward sharing of personal data
- measures taken to ensure Processors comply
- steps that have been taken to ensure storage limitation, and that suitable retention periods are applied to the processing
- how international transfers are safeguarded
- potential harms to individuals resulting from the processing or a breach of the personal data involved

- technical measures implemented to ensure confidentiality, integrity, and availability
- organisational measures implemented to ensure confidentiality, integrity, and availability

Following completion of the DPIA for each of the phases, the conclusion reached was that the development of the HDP posed no significant risks to data subjects. On this basis, the Programme successfully passed through each Gateway stage and the DPIA was signed off by the Jisc Data Protection Officer ahead of the Programme's production release.

The DPIA was supplemented by the output of Information Security Risk Assessments conducted at asset level by the Information Security SME.

Information Security Risk Assessments

The constituent systems within the Data Futures Programme were assessed with a standardised Information Security Risk Assessment which reviewed the confidentiality, integrity and availability controls within the system, assessed the impact of a breach of any of these factors and record and track any applicable risks. These assessments enabled the Programme to demonstrate that the processing of data within Data Futures had been investigated in a consistent manner to produce comparable results.

Data Futures, as a programme, was too large to be usefully captured within a single risk or impact assessment. As such, a hierarchal system of risk assessments was implemented to ensure high level breadth of coverage as well as, more importantly, specific detailed coverage.

TRANSITION (TO BUSINESS AS USUAL)

Ensuring Data Protection Principles continue to be met

The Check-in, Check-out approach supported the programme in identifying and embedding controls that ensured the design and build of the HESA Data Platform was done in a way that protects data subjects' rights and freedoms, not least against the following core principles:

Lawfulness, fairness and transparency

Articles 13 and 14 of the UK GDPR requires Controllers to provide privacy information to data subjects. This includes the Controller's purposes for processing personal data, including transfers and disclosures to other data Controllers. Jisc's DCS Collection Notices provide this information for students, staff and graduates on behalf of Jisc and the other legal entities who are Controllers of DCS datasets.

The Collection Notices are published at https://www.hesa.ac.uk/about/regulation/data-protection/notices.

HE providers are contractually required to inform students and staff that their personal data will be submitted to Jisc and that they must make the Jisc DCS Collection Notices available to all relevant data subjects. The DCS recommend that HE providers include a link from their own privacy notices to the Collection Notices.

In June 2022, ahead of the 2022/23 Student Collection, the Student Collection Notice was updated to accurately reflect the continuation of the Data Futures project (and the subsequent use of personal data) as the Programme moves into its first collection year using the HDP.

Purpose limitation

As previously described, the Data Protection SME worked at an Epic level with the Programme to adopt a Privacy by Design and Default approach from the outset. This ensured that every Epic was subject to an Information Security and Data Protection review at the outset, during and at the conclusion of the Epic.

This approach has supported the Programme in identifying and embedding controls that will ensure the design and build of the HDP is conducted in a way that protects the data subjects' rights and freedoms. Any new use of the personal data would be required to follow the same use process. During this, the Jisc Data Protection Officer would consider whether the newly proposed use of personal data is compatible with the purposes under which it has been collected.

Data minimisation

Every item of data collected is needed either by a Statutory Customer or Funder or to aid the collection process. Some data items are used in the derivation of datasets for Statutory Customers and Funders and then not further processed. The requirement for individual items of data is regularly reviewed as part of the record review process. Requests for new fields follow the DCS's Business Change Idea (BCI) process, during which they are Data Protection risk assessed.

Each data record is subject to a regular review and may be further amended to satisfy Statutory Customer or Funder initiatives in between planned reviews. Changes to records are mostly prompted by the needs of Statutory Customers or Funders or the desire to improve data quality and are subject to extensive consultation with all parties concerned, including the HE providers. Any changes must go through the Business Change Ideas process and are assessed from a Data Protection perspective. All the data protection principles are borne in mind during the record review process.

Individual identifiers

Collection of individual identifiers is essential both to aid the collection process and to allow the Statutory Customers and Funders to carry out their public functions effectively. These include the tracking of students across HE providers to produce accurate progression and participation statistics.

Student Names

Student names are needed to ensure the data collection process runs smoothly. Actual names are supplied to Statutory Customers for record linking and in support of audit processes. Names within the Student Record are not used to make direct contact with students. Access to names within Jisc and its Statutory Customers is restricted only to essential staff who have received training in data protection training.

In addition, the identification of individual students by those carrying out equal opportunity, research, journalism and other processing for statistical and research purposes is prohibited and, in the majority of cases, information is not shared on a named basis.

Students studying wholly outside the UK

Providers must not return individualised records for students studying for the whole of their course outside of the UK, or those not funded for study by distance learning overseas.

Accuracy

Jisc is reliant on the accuracy of data submitted by providers and therefore has recognised the risk that data processed and disseminated to third parties could potentially be inaccurate.

To improve data accuracy and to mitigate the risk of harm to data subjects, the following controls and processes have been implemented:

Quality rules and switches

The DCS has developed extensive quality assurance procedures and runs a range of automated validation checks (quality rules) against all submissions: https://www.hesa.ac.uk/collection/c19051/validation

Third party data used for validation and derived fields

Third party data is used to quality assure and validate submissions and to enhance the data collected through use of derived fields. A register is maintained to support each of these uses and any associated terms and conditions and required attributions for third party data is published on the HESA website: https://www.hesa.ac.uk/support/third-party-data

Storage limitation

An automated approach to managing the deletion and archiving of personal data has been implemented as part of the Programme. The records within the HDP are now retained based on the following schedules:

Information Asset Group Name	Action	Trigger	Proposed retention period	Automated / Manual	Legal / business justification
'Non-signed-off'* submissions and associated reports • Provider submission raw files • Provider submission data • Enriched provider submission files • Credibility data • Quality rule reports • Quality rule report download files • Additional collection reports **If the provider hasn't provided any signed-off data in the collection at the point it closes the latest submitted file (and associated reports) will be retained to support the Historic Amendment process.*	Delete	Collection state being updated to 'closed'	30 days after collection closes	Automated	The file submissions and associated quality reports for 'non-signed-off' transactions will be required to support the quality assurance process during the open collection as Jisc and providers need to be able to compare against these for each file submission for validation and quality assurance purposes. If files are deleted during the open collection, providers would not be able to benefit from the comparison functionality which has a potential impact on data accuracy / quality. Once the collection has closed, these non-signed-off files and associated reports are no longer required.
'Non-signed off' submissions and associated reports during historic amendment (fixed database in BAU) phase • Provider submission raw files • Provider submission data • Enriched provider submission files • Credibility data • Quality rule reports • Quality rule report download files • Additional collection reports	Delete	Collection state being updated to 'fixed database closed' (when historic amendment period closes).	30 days after the collection state has been updated to 'Fixed Database Closed'/the historic amendment period closes	Automated	The file submissions and associated quality reports for 'signed-off' transactions will be required to support the QA process during the historic amendment as Jisc and providers need to be able to compare against these to ensure only the approved changes have been made. Once the historic amendment period has ended, these reports are no longer required.
'Signed-off' submission files – • Provider submission raw files • Enriched provider submission files	Archive	Collection state being updated to 'fixed database closed' (when historic amendment period closes).	30 days after the collection state has been updated to 'Fixed Database Closed'/the historic amendment	Automated	The final 'signed-off' delivery files are required to be archived and retained indefinitely by Jisc for statistical research purposes.

Information Asset Group Name	Action	Trigger	Proposed retention period	Automated / Manual	Legal / business justification
			period closes		
Signed-off submissions data	Archive	Archive of the 'Signed-off submission files' for the next year's collection	30 days after the collection state for the next year's collection has been updated to 'Fixed Database Closed'/the collection state has been updated to 'Archived'.	Automated	Submission data will be used for historic data rules in the next collection and therefore it is required to be retained for an additional year beyond the typical historical amendment period. Jisc does not want to meet this data until the next year's collection has finished using it for the historic data rules.
'Signed-off' associated reports • Credibility data • Quality rule reports • Quality rule report download files • **Additional collection reports**	Delete	Collection state being updated to fixed database closed (when historic amendment period closes).	30 days after the collection state has been updated to 'Fixed Database Closed'/the collection state has been updated to 'Archived'.	Automated	Once the historic amendment period closes there are no identified use cases for retaining the collection reports associated with the final 'signed-off' delivery files.

Accountability

As a custodian for personal data in the HE sector, each phase of the Programme would not be approved for go-ahead via the Gateway Framework if the use of personal data was considered to present a risk to the rights and freedoms of data subjects, as defined by Recital 75 of GDPR.

Consequently, the approval by the DCS Data Protection Officer was provided at each release phase, following completion of the DPIA for the phase and only when satisfied that the processing of personal data as part of the build of the HDP did not present a high risk to the rights and freedoms of data subjects.

Integrity and confidentiality (security)

The technical and organisational security measures in place are outlined throughout the following pages.

Supporting Data Subject Rights Requests

Technical controls were implemented to support an approach that facilitates data subject rights, including the right to access, erasure, restriction and rectification. These technical capabilities have been implemented to fulfil rights requests received in Jisc's capacity as a Controller and will not be used to assist Providers or Statutory Customers in meeting their own requirements in this space.

It is important to recognise that the HDP is not intended to be a data repository for Providers and Statutory Customers. On this basis, Jisc would expect rights requests received by Providers and Statutory Customers to be fulfilled using data held in their own Controller systems.

TRAINING AND AWARENESS

Jisc staff and contractors

Jisc considers its employees to be a critical line of defence in protecting and securing the higher education sector and the data processed by Jisc.

Jisc's comprehensive training and awareness programme includes new employee onboarding, annual security and data protection training, role-based awareness education, and phishing simulations.

Jisc trains employees to identify often-used attack vectors such as phishing emails and how to report them. This applies to every employee and contractor.

In addition to training and awareness programmes, Jisc reviews and updates information security and data protection policies and procedures annually, and more frequently if needed.

Developers maintain contacts with specialist interest groups to keep up with industry best practice and emerging trends to keep an eye on the current Information Security threats and vulnerabilities.

Providers and statutory customers

The DCS Training team have delivered, and continue to deliver, a wide range of live training and e-learning content focused on Data Futures to Providers and Statutory Customers and funders.

Statutory Customers were kept informed of the programme by running a series of knowledge share sessions during Summer 2020. When delivering training on the Student and Student Alternative records in subsequent years, the Training team have continued to acknowledge the Data Futures programme to prepare delegates for the transition to the new record.

In 2022, the Training team designed and released a suite of provider-focused e-learning courses while also delivering a series of live webinars to support the introduction of the HDP. As part of the Alpha pilot, the team released an introductory course specifically for Alpha participants. This was followed by the popular ***This is Data Futures*** course which was designed to introduce the sector to the Data Futures programme in an accessible manner and launch the new portfolio of courses.

Between March and June 2022, the Training team delivered eleven live webinars to the sector. These provided all delegates with a comprehensive introduction and walkthrough of the record, while drawing parallels and comparisons between their legacy record and the new Student model. In total, the team trained over 350 delegates.

LOCATIONS OF PROCESSING

The development and deployment of the HDP to the cloud introduced new requirements and approaches to security and data protection, not least the need to consider and risk assess any international transfers of personal data.

Underlying Infrastructure

The HDP utilises cloud services from Amazon Web Services (AWS) for data submission, processing, and delivery.

The AWS servers used to service HDP are located within the United Kingdom using the eu-west-2 availability zone. An additional zone is located within Ireland for redundancy purposes and further redundancy regions are available within Europe should the requirement arise.

An underlying component that does not process personal data is hosted out of the AWS Global region due to the increased functionality available.

Supporting Infrastructure

The HDP processing will be supported by internal systems and external named delivery partners in the following regions:

System	Description	System Owner	Personal Data Processing	Cloud Service	Processing Location
Issue Management System (IMS)	Used for Issue Management within HDP	Jisc	Yes	Azure	United Kingdom EEA
Identity System (IDS)	Used for authentication within HDP	Jisc	Yes	Azure	United Kingdom EEA
Reference Data Store (RDS)	Holds third party reference data used in the quality assurance process	Jisc	No	Azure	United Kingdom EEA
Personal Data Store (PDS)	Holds third party reference data used in the quality assurance process	Jisc	Yes	Azure	United Kingdom EEA
Help Scout	Used to liaise with Providers and Statutory Customers	Help Scout PBC	Yes	N/A	United States
Salesforce	Service used to store contract details for users of DCS systems in order to issue system notifications	Salesforce.com, inc.	Yes	N/A	EEA

System	Description	System Owner	Personal Data Processing	Cloud Service	Processing Location
SendGrid	Used to send email notifications to Providers and Statutory Customers regarding their HDP accounts and role management. Used by IMS to notify users of updates to issues	Twilio Ireland Limited	Yes	N/A	United States
Cloudflare	Provides DDOS and WAF capabilities to Jisc's web estate	Cloudflare, inc.	Yes	N/A	Processes data in the data centre closest to the end user. This means that any users outside the UK or EU will likely have data processed in their country

With the exceptions of Help Scout, Salesforce and SendGrid, the systems and components listed that comprise the HDP are hosted in Jisc's tenants within AWS and Microsoft Azure (hosted in the UK and European Economic Area).

Following advancements in the area of international transfers, Jisc have undertaken risk assessments of its transfers of personal data outside of the United Kingdom and EEA in order to ensure adequate safeguards are in place with key delivery partners.

To this end, Jisc has reviewed the existing contractual arrangements in place with Help Scout and SendGrid. The existing agreements with both third parties incorporate the necessary Standard Contractual Clauses and International Data Transfer Agreements.

DATA CENTER SECURITY

Amazon Web Services (AWS) is a best-in-class enterprise cloud computing platform which implements a comprehensive host of security controls up to Government level. AWS is accredited to ISO/IEC 27001:2013, 27017:2015, 27018:2019, 27701:2019, 22301:2019, 9001:2015, and CSA STAR CCM v3.0.1 security standards.

More information regarding AWS security controls can be found using the following link: https://aws.amazon.com/compliance/data-center/controls/

HOST AND NETWORK SECURITY

The HDP uses an industry best practice level of encryption to ensure that all data within the Programme is secure at all stages of the data journey. All the services, such as databases, are encrypted at-rest using the industry standard AES256 encryption algorithm. All communication that is in transit across the HDP is encrypted using TLS v1.2 and v1.3 ciphers. Communication paths from an end user are encrypted all the way through to the back end of the system.

PENETRATION (PEN) TESTING

The DCS maintained a programme-specific Penetration Testing Roadmap.

This Penetration Testing Roadmap ensured that development work was tested at defined stages before release into Alpha, Beta or Production. Testing was undertaken by the DCS's external independent partner, Bridewell.

A total of 8 system wide penetration tests were conducted during the 6-month Alpha testing phase, the 10-month Beta testing phase and prior to the Production release. This Penetration Testing Roadmap was structured to ensure it aligned with the secure delivery of significant functionality.

As part of the testing process every finding was prioritised according to its severity. After each test, all critical and high findings identified required remediation before the Gateway Framework gave approval for the code to be deployed. Lower priority findings were triaged with remediation plans and deadlines agreed.

Successful completion of a relevant penetration test and remediation of any critical and high findings were a mandatory requirement before the Programme could pass beyond the Gateway phase.

Sign off process

Following the publication of the independent external Penetration test report, initial triage meetings and internal meetings were held with project and business stakeholders to agree target deadlines and outline remediation plans for each finding. Some findings that were of a low priority with existing mitigations were logged within the programme's risk register and accepted for ongoing monitoring.

Remediations to findings were assigned to relevant development teams, built and verified by the DCS Information Security function before implementing into the release and any residual risks were mitigated.

When Data Futures moves into business as usual (BAU) it will fall under Jisc's existing Technical Vulnerability Management Policy which describes the automated vulnerability scanning and regular penetration testing processes.

INFRASTRUCTURE LOGGING AND MONITORING

The Data Futures programme has implemented full logging and monitoring within the DCS Security Information Event Monitoring (SIEM) system, Microsoft Sentinel, allowing for potential events and threats to be managed in real-time.

The DCS undertook threat modelling scenarios to identify the use casing for developing a suite of monitoring alerts, which were then integrated into the SIEM and monitored using standard operating procedures. The system generates alerts based on correlated detection logic and notifies the Jisc incident response teams who then investigate the causes of the alerts using standard processes and procedures.

In addition, Jisc has developed an API Design Policy which, in conjunction with the Secure Development Policy, mandates the requirement for system auditing and logging at the point of system design.

USER ACCESS, IDENTIFICATION AND AUTHENTICATION

HESA's Identity System (IDS)

The DCS Identity System (IDS) enables users to have a single account to access our Data Collection System, Issue Management System data quality database, Graduate Outcomes portal and (or) Heidi Plus.

Roles which govern the level of user access are required within IDS to access DCS systems. Separate roles are required for each system. A list of the various roles and their responsibilities can be found in the IDS user guide.

For each data stream, the Record Contact at a provider is the first point of communication during data collection. The Record Contact is responsible for overseeing a provider's data submission process and ensuring that deadlines are met. Record Contacts are administered by the DCS.

Record Contacts can administer the Data Collection roles themselves. If administering roles themselves, Record Contacts will be expected to invite people to hold roles and to respond to colleagues' requests for roles.

Record Contacts are responsible for gatekeeping the platform by ensuring that people who either no longer act for an organisation or no longer have a role in the submission of data, have their roles revoked.

The DCS Liaison team run an annual review process for the grantable roles associated with each collection. This requires the Record Contacts at providers to review the roles held for their collections and confirm if they are still appropriate or to revoke the roles.

Multi-Factor Authentication (MFA)

Users of the Identity System (IDS) and any of the systems it governs are required to set Multi-Factor Authentication on their accounts to gain access. The DCS provides guidance on registering an account and setting up MFA to support providers and maintain the security of the systems.

RISK MANAGEMENT

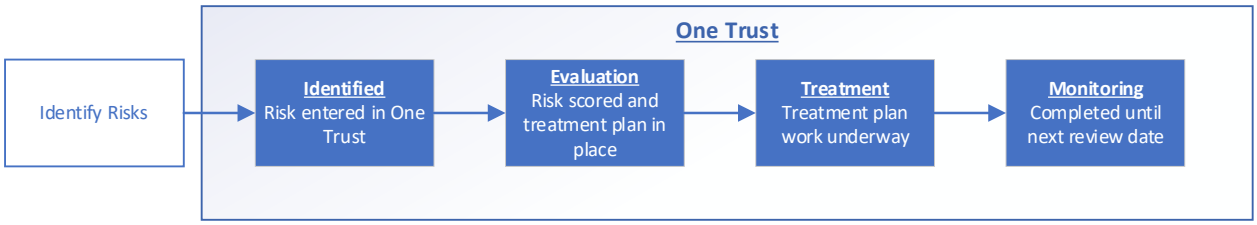
The Gateway Framework described earlier in this document ensured that the necessary Data Protection and Information Security Risk Assessments had been undertaken ahead of Programme milestones as well as on an ad-hoc basis, where appropriate. This was in addition and supplementary to the completion of mandatory DPIA, as described above.

These assessments enabled the project to demonstrate that the processing of data within the Programme had been investigated to identify applicable Information Security and Data Protection controls, to mitigate potential risks to the rights and freedoms of individuals, as well as to safeguard the confidentiality, integrity, and availability of the data.

To mitigate identified risks, a risk management methodology aligned with ISO 27001:2022 and ISO 27701:2019 was followed. The DCS uses GRC and Privacy Software OneTrust to support its approach to risk management.

Risk assessment process

OneTrust is used to record and manage risks, following the four-step process below:



Identification

The risk is identified and entered into the OneTrust system capturing relevant information such as the threat, vulnerability, risk scenario and risk treatment.

Evaluation

During this phase, a plan to treat the risk to reduce the Net (residual) risk to meet the Target risk level is produced. Depending on the Net risk score it is decided whether a treatment plan should be added or if no further controls are required.

Treatment

Tasks identified in the treatment plan are completed. As this progresses the risk owner ensures that the risk is reviewed and that the Net score is amended as appropriate. The risk will remain in the “Treatment” phase until all the required treatment measures are completed.

Monitoring

This is the steady state that most risks once reduced to an acceptable level are in. Risks which meet the Target risk score are reviewed on at least an annual basis by the risk owner to see if

anything has changed. Risks which do not meet acceptable levels (Target risk score) are reviewed regularly to assess if there are any viable measures to reduce risk further.

SECURE DEVELOPMENT LIFECYCLE

The development teams working within the Programme followed a secure development lifecycle. This is being completed using agile development methodologies via 2-week sprint cycle. The lifecycle included the following phases:

- Requirements
- Architecture and design
- Coding
- Testing
- Release and maintenance

As part of embedding information security and data protection principles into the development of the HDP and supporting systems, considerations are made at each phase.

Requirements

Specific requirements for Information Security and Data Protection are captured during this phase and refined with the SMEs.

Architecture and design

The ‘Check-in’ step as detailed previously in the paper, takes place during this phase.

Coding

The team work to a Secure Development Policy.

Testing

Tooling, including Snyk and Sonar Cube, are used during this phase to test the code developed during the sprint.

Review

The ‘Check-out’ step as detailed previously in the paper, takes place during this phase.

Release and maintenance

During the Alpha and Beta phases a Penetration test was completed before each software release. The Programme completed its last main Penetration test as part of preparations for Go-Live. The teams will move into a Continuous Improvement, Continuous Development (CICD) approach, akin to the processes within business as usual. The approach has been approved by the relevant SMEs to confirm that the right considerations are being taken, to ensure compliance with relevant information security and data protection principles and requirements.

INCIDENT RESPONSE

Jisc have an established information security and personal data incident process which encompassed the Programme and will continue to do so following the transition to business as usual.

In the event of a near miss, event, or incident, the Information Security and Data Protection teams follow a seven-step process which includes:

1. New - A new Incident is logged.
2. Investigating - The nature and scope of the incident is understood with the assistance of relevant personnel. Containment and remedial action to be taken is agreed.
3. Containment - The containment actions identified during the investigating are completed and recorded.
4. Notifying - A decision is made by the Data Protection Officer, or Deputy, on whether the incident is in fact a personal data breach requiring notification to the Supervisory Authority, data subjects or any other interested parties.
5. Lessons learned - Once the initial incident is contained and recorded, a full review of the causes of the incident; the effectiveness of the response(s) and whether any changes to systems, policies and procedures is undertaken.
6. Review – A review is undertaken and approved before sign off is completed.
7. Complete – Once all steps identified in management of the incident have been completed, the incident is closed. Actions resulting from lessons learned remain open as sub-tasks and are regularly reviewed.

BUSINESS CONTINUITY AND DISASTER RECOVERY PROTOCOLS

Data backup schedules are configured within AWS with back up vaults implemented in the eu-west-2 and eu-west-1 AWS regions. Snapshots are taken daily at midnight for databases and S3 buckets containing file uploads and deliveries.

Continuous backups are also enabled in the AWS Backup Vault and continuous backups are taken for RDS and the file uploads and deliveries S3 buckets. Continuous backups run every 15 mins into the primary vault (eu-west-2).

AWS Backup is an AWS managed service with high quotas to ensure scalability as the volume of data processed by the HDP increases. See https://docs.aws.amazon.com/aws-backup/latest/devguide/aws-backup-limits.html

AWS Backup retains backups in the primary eu-west-2 vault for 35 days. The midnight snapshots are copied into the secondary daily and retained for 1 year.

The Disaster Recovery process was tested as part of the staging deployment phase of the programme. Going forward, Disaster Recovery procedures will be tested on an annual basis.

JISC PARTNERS WITH PROVIDERS

Throughout the development of the HDP, Jisc has continued to support the HE sector in the following areas:

Onboarding

When a new HE provider is being onboarded, they must sign Jisc Subscription Agreement before gaining access to the IDS system.

Training

Jisc provide a variety of live, web-based and bespoke training opportunities, helping HE providers respond to their statutory data requirements.

Jisc training is designed to help to develop in-house expertise on all aspects of the Collection data journey and so training on data collection and submission is delivered.

See the Training and awareness section for details on how the Training team have delivered Data Futures Programme training.

Guidance

The HESA website offers a comprehensive resource of help and guidance to providers. User guides are available on topics from the Issue Management System (IMS), the DCS's data quality database, to how to use PivotTables. Technical guidance on submitting data to Jisc is available in the relevant coding manuals.

Multi-Factor Authentication (MFA)

Users of IDS and any of the systems it governs are required to set Multi-Factor Authentication on their accounts to gain access. Jisc provides guidance on registering an account and setting up MFA to support providers and maintain security of the systems.

Data protection guidance

A wide range of data protection support is available on the Jisc DCS (HESA) website. Providers can access collection notices for the Student, Staff and Graduate Outcomes records, which describe the purposes for which the data is collected.

Data protection guidance notes are included in each release of the data collection coding manuals. These provide information about how data protection legislation affects the processing of data by Jisc (i.e., the Student record data protection guidance).

Information is also available for HE Providers undergoing a merger, to ensure that receiving and sharing personal data with merged entities is in compliance with data protection laws and the Data Sharing Code.

Cyber security for providers

Cyber-attacks on providers raise the risk of having an impact on the integrity of Jisc's systems and data. Jisc has a standard procedure for responding to cyber-attacks, and early notification to Jisc means it can respond quickly to safeguard data submitted to the HDP.

Communications

HE Providers have a range of channels open to them to engage with Jisc. Four email newsletters are published regularly covering data collection and Jisc news, open data, Data Futures and training updates.

Two groups and forums meet with representatives from across the sector:

- Provider forum – for knowledge exchange between Jisc and sector representatives to develop services to meet the combined needs of the HE sector.
- Graduate Outcomes steering group – provides transparent advice to Jisc on the implementation and administration of the Graduate Outcomes survey.

Jisc runs consultations to monitor and engage with a variety of stakeholders to maintain a comprehensive understanding of the changing HE landscape. Several Data Futures consultations were run during the Programme including readiness surveys and Alpha and Beta expressions of interest.

Annex 1 – Privacy and Security Requirements Checklist

The Data Privacy and Information Security requirements were specific to the individual item for each Epic. To help identify relevant requirements for each Epic, a standard 'checklist' of questions was formulated with the intention that these are applied to each backlog item at the outset and revisited periodically throughout the item's lifecycle.

A template of the checklist is given below:

Check	Description	Response
What confidential information and personal data is this work going to be processing / facilitating / creating / reading / updating / deleting?	Confidential information is information that has value to HESA, may amount to a trade secret, should not be disclosed publicly. Personal data is data that directly or indirectly identifies an individual (including identifiers such as HUSID) and is split into general personal data (such as name, date of birth), special category data (such as health, disability, ethnicity, religion, sexual orientation) and criminal offence data .	Identification of information types (e.g. collection data, staff data, personal data, IP information, technical data such as login data etc.) for further investigation with the DF Compliance team.
How will **data minimisation** be ensured so that only personal data that is adequate, relevant and limited to what is necessary is processed?	Data minimisation means only processing personal data that is necessary for the purpose.	Justifying why the personal data is being processed. Consider all processing to be applied in this backlog item (e.g. all processing from the intake of the personal data through to its destruction).
How will data be kept **accurate** and up to date in the processing?	Personal data must be kept accurate and up to date and erased/ anonymised without delay. This also relates to recording information assets and retention periods (which are maintained in the Data Futures Information Asset Register).	Is the personal data able to be amended or updated if necessary? Has the Information Asset Register been updated to reflect any new information assets and retention periods agreed?
How long will personal data need to be retained for and will it be anonymised or deleted at the end of its retention period?	Personal data can only be retained for as long as is necessary. The retention must be justified. This information is kept centrally in the Information Asset Register.	See above.
How will we ensure the **confidentiality** of this information?	Sensitive and confidential information must only be available to those who are authorised to see it.	List of controls to be implemented to ensure confidentiality is maintained. Such as access controls, data segregation etc.
How will we ensure the **integrity** of this information?	The consistency, accuracy and trustworthiness of the information must be maintained.	List of controls to be implemented to ensure integrity is maintained. Such as encryption at rest and in transit, auditing etc.
How will we ensure the **availability** of this information?	The information must be consistently and readily available to authorised parties.	List of controls to be implemented to ensure availability is maintained. Such as continuity plans, capacity management, UI designs.
What systems are involved in this work?	Any system that acts as a touchpoint for information must be assessed for potential risks.	List of systems that store, process, display and transmit sensitive and confidential information.

Check	Description	Response
What information security risk assessments are required for these systems?	Any system that has not been risk assessed for its intended purpose, or any system that is developed, significantly changed, or pre-existing and brought into Data Futures, will need to be risk assessed for the current work package.	List of applicable systems from the above list of all systems.
Have third party software or solutions been assessed for legal, information security and Data Protection risks and issues?	The process of onboarding third party solutions must be followed to check licence provisions, data protection considerations (such as international transfers and Processor provisions), and information security risks have been considered.	Confirmation of process followed to onboard and incorporate third party solution.
Are any new development methods, technologies, standards or practices being used in this epic?	All development work must adhere to the agreed secure development policies. Anything different must be assessed to ensure it complies with these policies and are documented.	Confirmation that any new development methods, technologies, standards or practices used within the Backlog item is in alignment with HESA's Secure Development Policies.
What documentation will need to be produced?	Necessary documentation detailing the completed product, its use and its interactions etc. will be required, for example Confluence page updated, Operational Documentation/ User Guide created, Code commenting etc.	Identification of relevant documents that will need to be produced for the item and identification of authors.
What user training is required?	Users of the completed product must have the knowledge to be able to use the product correctly to maintain the security of information.	Identification of required user training and training providers, where relevant.
Will any penetration tests be required?	All developed systems that are intended to be externally facing and could potentially expose sensitive and confidential information, will need to undergo an external penetration test and subsequent vulnerabilities will need to be remediated.	List of applicable penetration tests. Precise scopes will be determined closer to the time of the test.
Has the Compliance Team been consulted about whether a DPIA needs to be created or updated?	Any processing involving personal data needs to be impact assessed and consideration about whether a Data Protection Impact Assessment needs to be completed or updated. The purpose of a DPIA is to ensure HESA is appropriately identifying and mitigating risks to the rights and freedoms of individuals that may be posed by the processing being developed.	Consult with the DF Data Protection Compliance Officer if the processing involves processing of personal data (other than incidental data, for example HESA staff data).

Contact details
HESA
95 Promenade
Cheltenham
GL50 1HZ
E data.protection@hesa.ac.uk
T +44 (0)1242 388 513 (option 5)
W www.hesa.ac.uk