Electronic Data Retention, Archiving and Purging Policy

LSE holds a large amount of data in databases, on file systems and elsewhere to support the operational business of the School. The volume of this data is increasing as new applications and services are implemented and as additional data is required or becomes available for use. This continued growth in volume is unsustainable without a corresponding commitment to data management activities to identify and remove data that is no longer relevant or need not be retained.

Retaining data for which there is no longer a need incurs both a financial and environmental cost and reduces the operational efficiency of School systems and processes. It increases the risk and impact of data breaches and may also mean that the School is not meeting its legal obligations.

This document sets out the School's Policy on the retention and purging of electronic data, and provides guidance for data archiving.

Purpose

The primary purposes of this Policy are to:

- Clearly define what is meant by retention, archiving and purging of data.
- Set out the principles and requirements for data to be purged.
- Outline the responsibilities for purging of data.

This Policy does not attempt to describe any of the technical processes required to undertake the purging of data, which will vary across systems and scenarios.

Scope

This Policy covers all types of electronic data and is not limited only to personally identifiable information (PII) or otherwise sensitive data.

This Policy applies to:

- Data held in LSE-owned databases, applications and systems, irrespective of the physical storage location.
- Data held in LSE instances of applications provided on a Software as a Service (SaaS) or Platform as a Service (PaaS) basis, e.g. Salesforce and OneFinance.
- Data held in file servers, network shares and cloud storage 'buckets'.
- Structured and unstructured data held in the Microsoft 365 environment (including but not limited to Teams, SharePoint and OneDrive).
- Exports, extracts or copies of data taken from its original location and held in a database, spreadsheet or other file format, for whatever purpose.
- Backups of data, taken either as part of a formal backup process or informally for operational reasons.
- Data created by individuals and machine-created data.

This Policy does not apply to:

- Data created, obtained, or used for academic research purposes (data used for the management of research, however, is in scope).
- Data not held in an electronic format e.g. paper records[1].

Accompanying documents

This Policy is supported by the following documents:

Data principles and Non-live data principles that provide the architectural approach specifically pertaining to data in School environments.

---

[1] Paper records are covered by the Destruction of Physical Records procedure.

The Data Governance Framework and Operating Model that defines the roles and responsibilities related to the ownership of data.

The School Retention Schedule that provides information about the time periods for which records of particular types are required to be retained, after which point records must be either archived or destroyed.

Procedure for Database Purging that sets out the process for making decisions about data to purge and how to communicate and record that data has been purged.

The Log Retention Policy that sets out sets out the policy on the creation and retention of logs produced by LSE applications, systems and infrastructure.

Definitions

Retention

According to the Data Management Association (DAMA)[2], retention refers to:

How long data is kept available.

Retention of data is determined by business requirements, both operational and strategic, whilst also considering data security and risk, and abiding by any legal or regulatory obligations.

Retention is directly related to archiving and purging. As the need for data to be available for use changes it may be appropriate for that data to be archived. When there is no longer a need to retain the data at all then it must be purged.

As an example of retention, the School has a legal requirement to retain financial data for seven years but additionally has a business requirement to retain that data for a longer period of time. When the legal and business requirements no longer apply then the School must purge that financial data.

Archiving

The DAMA definition of data archiving is:

The process of moving data off immediately accessible storage media and onto media with lower retrieval performance. Archives can be restored to the originating system for short-term use.

When data no longer needs to be available for immediate use but otherwise does need to be retained to meet a business, legal, or regulatory requirement, then it can be archived. Typically this means that the data is moved from the primary system or database where it was being used and maintained, to a separate system or database designed and provisioned for archiving

---

[2] Data Management Association (DAMA) International Guide to the Data Management Body of Knowledge (DAMA-DMBOK2), second edition, 2017

purposes. Generally an archive system is less accessible than the primary system, has lower levels of performance, and uses cheaper storage.

It should be possible to restore archived data back to the originating system or database on a temporary or permanent basis if the need for this arises.

A key aspect of archiving data is that the data remains available to the School.

Purging

DAMA provides the following definition of purging:

The process of completely removing data from storage media such that it cannot be recovered. Data to be purged is generally deemed obsolete and unnecessary, even for regulatory purposes.

When there is no longer a business, legal, or regulatory requirement to keep the data – either in an archive or originating system – then it must be purged. After purging it is no longer available to access or use and cannot be recovered or restored.

Returning to the earlier example, the School must purge financial data after seven years and there is no remaining business justification to keep it for a longer period of time. The data will be purged regardless of whether it has been archived or is still held in the Finance system.

Purging may also be referred to as removal, deletion, or destruction but note that the operation must be permanent to be considered purging.

A consequence of purging data is that the data is no longer available to the School.

Dataset

For the purposes of this Policy the term 'dataset' will be used to refer to any collection of data that is subject to retention, archiving or purging, and applies to all instances of that collection of data regardless of format and location. For example, a dataset might consist of student examination results for a specific academic session and that dataset may exist in multiple formats and locations:

- in the Student Records System and corresponding database
- in the testing and development instances of that system and database
- in formal backups of that database that are held in Cloud-based storage
- exported into a spreadsheet kept in OneDrive.

If that dataset is purged from the Student Records System and corresponding database, then it must also be purged from the other locations listed. It should be clear that to abide by this Policy it is essential that the location and use of School data is fully understood.

Policy

This Policy sets out the principles and requirements for data retention, archiving and purging. It

does not cover any scenario in detail, but it does provide a high-level set of principles that must be adopted and followed by those responsible for the implementation of purging procedures.

Retention

Every dataset must have a documented retention schedule that states the time period after which the data will be purged, and if applicable the conditions under which data will be archived. The School Retention Schedule provides information to assist in the creation of individual retention schedules, and a template document is available from the DTS Strategy and Architecture Team.

It is the responsibility of data owners or information asset owners to ensure that retention schedules are defined and followed. Coverage of retention schedules will be reviewed by the Data Governance Group[3] and overseen by the Data, Technology, and Digital Management Board.

Bear in mind that the retention of data does not necessarily mean that all the data needs to reside together in the same system or database, with the same level or speed of access. Data can be archived and remain accessible for use.

Archiving

When a dataset no longer needs to be as immediately available as other related data (for example, data relating to completed projects compared to projects that are still in progress) but there remains a legitimate business, legal, or regulatory requirement to retain that dataset then it should be considered for archiving.

Many School systems do not provide a technical mechanism to allow the archiving of data. If a technical mechanism is provided then archiving is mandatory. See Appendix A for guidance on archiving.

Archived data remains subject to its retention schedule. Archiving is not the end state for a dataset. Based on the conditions and/or time period defined in the retention schedule the dataset will at some point be purged. Archiving, therefore, is not equivalent to or an alternative to purging data.

Purging

The retention schedule that applies to the dataset will set out the time period after which the data will be purged. If there is no longer a legitimate business, legal, or regulatory requirement to retain the dataset then it must be purged according to the retention schedule.

It is the responsibility of data owners or information asset owners to ensure that the dataset is purged. As stated before, this applies to all instances of that dataset regardless of location or format. Purging activities be reported to the Data Governance Group and overseen by the Data,

---

[3] The Data Governance Group will be established in academic year 2022/23.

Technology, and Digital Management Board.

Purging must be a regular activity, performed either periodically or based on specific conditions being met as defined by the retention schedule.

It is important to reiterate that purging data means it is permanently removed, is no longer available to access or use, and cannot be recovered or restored.

Responsibilities

All LSE members

All LSE members are responsible for meeting this Policy through their use of School data.

Data Owners or **Information Asset Owners**

Data owners or information asset owners are responsible for defining the retention schedules for their data, and for ensuring that those retention schedules are followed as described in this Policy. They may choose to delegate that responsibility, but they remain accountable for the retention, archiving and purging of their data.

Data and Technology Services (DTS) and **Business Led Technology Teams (BLTs)**

Staff in DTS and BLTs are the custodians of data and will undertake technical activities, as directed by the data owners or information asset owners and in accordance with the defined retention schedules, to archive and purge data from systems, databases and other storage locations.

Architecture Board

Responsible for reviewing and endorsing this Policy.

Information Governance Management Board (IGMB)

Acts as a forum to consider any issues relating to this Policy and its application in conjunction with related Policies, procedures and guidance.

Data, Technology, and Digital Management Board (DTDMB)

Responsible for approving this Policy. Ensures that data retention, archiving and purging is factored into the operational plans of BLTs. Provides oversight and assurance that this Policy is followed and an escalation route for issues relating to adherence.

Appendix A – Guidance on Data Archiving

When a dataset no longer needs to be as immediately available as other related data but there remains a legitimate business, legal, or regulatory requirement to retain that dataset then it should be considered for archiving.

In addition to the definition of archiving provided in this Policy it should be noted that data archiving is different from and separate to the institutional LSE Archive[4].

Archiving systems generally have lower levels of performance and make use of cheaper less available storage infrastructure than that used for the primary systems.

Typically data that has been archived is still required to be used for reporting and analysis purposes. There are implications for the methods used to access archived data that will need to be considered when archiving is implemented.

Effective archiving should allow for an archived dataset, or a subset of it, to be restored to the originating system or database either temporarily or permanently. This will depend on the technical capabilities of the system and the way that archiving is implemented in each case. The restoration of archived data to the originating system must be periodically tested and validated.

Archiving is not a one-off activity. Archiving must be performed on a regular basis, either periodically or based on specified conditions being met. Details of these conditions and the frequency of archiving must be documented in the retention schedule.

Archived data remains subject to its retention schedule. Based on the conditions and/or time period defined in the retention schedule the dataset will at some point be purged.

Many School systems do not provide a technical mechanism to allow the archiving of data. If a technical mechanism is provided then archiving is mandatory. Advice and assistance with data archiving can be provided by DTS.

---

[4] More information about the LSE Archive is available here: https://info.lse.ac.uk/staff/divisions/Risk-and-Compliance-Unit/LSE-Archives

Review schedule

Review interval	Next review due by	Next review start
Annual	January 2023	November 2022

Version history

Version	Date	Approved by	Notes
1.0a	17 November 2021	Architecture Board	
1.0b	12 January 2021		Added reference to retention schedule templates; reformatted
1.0c	24 January 2021	DTDMB	

Links

Reference	Link
Data principles	https://lsecloud.sharepoint.com/:b:/r/sites/imt/Policy%20Book/Architecture/Data%20Principles%20v2.pdf?csf=1&web=1&e=oajfJw
Non-live data principles	https://lsecloud.sharepoint.com/:b:/r/sites/imt/Policy%20Book/Architecture/Non-Live%20Data%20Principles%20v2.pdf?csf=1&web=1&e=x7BIzg
Data Governance Framework and Operating Model	https://lsecloud.sharepoint.com/sites/TEAM_LSE-Data-Strategy-and-Governance/Shared Documents/General/Approval Pack/LSE Data Governance Operating Model.pdf
Destruction of Physical Records procedure	https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/desPro.pdf
Retention Schedule	https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/retSch.pdf
Procedure – Database Purging	https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/internal/datPurPro.pdf
Log Retention Policy	https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/internal/logRetPol.pdf

Contacts

Position	Name	Email	Notes
Data Architect	Steve Earley	s.earley@lse.ac.uk	Author
Information and Records Manager	Rachael Maguire	r.e.maguire@lse.ac.uk	

Communications and Training

Will this document be publicised through Internal Communications?	**Yes**
Will training needs arise from this policy	**Yes**
Communication about the publication and use of this policy will be done via:	

1. Notification to DTS via Teams
2. Communication at DTS Service Leaders' Group
3. Notification to Business Led Technology Teams via Teams
4. Notification to the Business Improvement Unit via Teams or email
5. Communication at the Business Led Technology Teams Ops Forum
6. Email notification to any other interested parties, not covered by these groups
7. Updates, eg termly, at DTS All Staff meetings on policy writing, including all new policies