Guidance on Retention and Disposition of Personally Identified Information in University Data

University data can be a valuable resource to facilitate data-informed decision making, document institutional history, advance academic research and scholarship, and support Yale's academic mission. However, not all university data is equally valuable and retaining some data beyond its business purpose can pose a risk to the university and to the privacy of any individuals referenced in the data. University data should only be retained indefinitely where there is a reasonable expectation of value. For example, records of institutional governance decisions or research data that is valuable for future research use may need to be retained indefinitely.

The need to retain university data should be balanced against risks of retaining the data. Concerns arising from retaining data include:

- Exposure of information in a cyber-attack or other data breach leading to regulatory scrutiny, reputational damage, and incurring costs of fines, notice, and mitigation;
- Financial costs arising from storage fees, maintenance of secure storage systems, and costs of processing data subject rights requests;
- Liabilities arising from breach of contract where university data is retained beyond contractual limits;
- Reputational damage associated with poor data management practices;
- Potential for using inaccurate or outdated information in decision making.

For example, retention of sensitive data—such as social security numbers—poses a privacy risk in the event of a breach that could lead to mandated reporting to data subjects, state and federal authorities and associated fines, mitigation costs, and reputational harm. Redacting social security numbers from a data set that otherwise merits retention would help to minimize these risks.

This guidance provides considerations to assist University Data Stewards in developing data retention and destruction practices for any university data with personally identifiable information under their purview. Data Stewards are encouraged to consider the data in their domain and document retention and disposition expectations for that data. Documentation should describe the applicable data or classes of data in the records, define the minimum retention period, state expectations for disposition of the data following the minimum retention period, and assign responsibility for ensuring university data is managed in accordance with the document.

High and moderate risk data are required to be purged once it has met its retention period under Yale's Minimum Security Standard 7.7[i]. Privacy by Design[ii] principles recommend considering data disposition at the time of collection and should be addressed during the Security Planning Assessment [iii](SPA). Effective Privacy by Design practices prior to data collection can ease end of data life management through implementation of data tags or creating database structures that facilitate identifying data fields to be purged at the end of the retention period. Developing thoughtful data collection and storage practices initially will facilitate data management and

reduce the risk of having co-mingled data increasing the difficulty in purging data or parts of data sets that merit destruction.

Development of Data Retention and Destruction Plans:

1. **Defining Applicable University Data**
   University data includes all records created or received by Yale faculty, staff, students, trainees, volunteers, contractors, or agents while acting on behalf of Yale, as well as data created or received by Yale students or trainees while providing a service to Yale or to others as part of their education or training. University data does not include intellectual property which by law or by Yale's copyright or other policies is owned, licensed, or otherwise legally controlled by the intellectual property creator. University Data Stewards are responsible for all university data under their domain as described in Policy1601 *Information Access and Security* and https://yaledata.yale.edu/about/data-governance-sponsors-data-stewards?app=yalesites and should develop an overarching retention schedule for the data under their purview.

2. **Defining Minimum Retention Period**
   University data must be retained for the period of time necessary to meet the operational, administrative, and legal requirements of the University. Stewards should consider any regulatory (https://ogc.yale.edu/sites/default/files/files/Yale-Records-Retention-Schedule(1).xls) as well as contractual obligations associated with the data to determine if there are minimum retention obligations that must be meet when determining an appropriate retention period. In addition, there may be business practices that require access to university data beyond what is legally or contractually required. Stewards should consult with data users to develop a retention schedule that reflects institutional needs.

3. **Disposition of data**

   a. *Archiving Data*: At the end of the data retention period, Data Managers and data users who believe there is an ongoing University need to retain the data, including but not limited to archival value, historical significance, or ongoing use of the record, should discuss the need to extend the retention period with the Data Steward. The Data Steward should consider what data or portion of data sets would be retained, the proposed end date of the extended retention period, if any, and the justification/rationale for the request. Consideration should be given to the classification of the data, with high-risk data being held to a higher standard to justify retention. Stewards should consider if any high-risk data elements can be redacted, if the data can be pseudonymized, or if the data can be retained in an aggregate form while still meeting the goal for archiving the records, thereby only maintaining the minimum necessary information. For example, regarding data that includes social security numbers, one should consider if the numbers can be removed prior to archiving without materially impacting the usefulness of the data set. In situations where the information is needed for year over year trend

analysis, the data may be able to be aggregated into groupings that would reduce the risk of identifying a given individual or data point in the records. Ultimately, decisions to retain university data beyond the retention period need to consider the risks associated with retaining the data against the actual likelihood of the data being of value in the future. Data that has not been accessed for several years are likely to have a low future utility.

Where there is a legitimate need to retain the data, a copy of the data may be retained either in an active state in existing systems or archived in a secured data repository in accordance with Yale's Minimum Security Standards[iv]. Where there is disagreement with the Data Steward's decision regarding disposition of a data set, the matter may be escalated to the Data Governance Executive Council for review. When data is to be archived, a single copy of the data would be retained while all other copies of the data in all formats should be destroyed. In some cases, the Department of Manuscripts and Archives may determine that the data is of permeant institutional value and will arrange for appropriate storage.

Archived data should include documentation of the rationale for retention and the new destruction date, if any, and will require an approved exception request[v] to retain moderate or high-risk data beyond its retention period.

b. *Data Destruction*: Subsequent to the minimum retention period, university data should be destroyed unless there is a legitimate business need to retain the information as described above and in accordance with Yale's Minimum Security Standards (see MSS 7.7). Without a determination that the data merits retention, all copies of the data in all formats must be destroyed including paper records and electronic files (e-mail, written documents, spreadsheets, databases, materials in imaging systems, etc.).

Data should be destroyed in accordance with Policy 1609: Media Control. Where appropriate, Data Stewards should consult with any known users of the records as well as the Office of the General Counsel prior to destruction to confirm that there is no reason to extend the retention period based on litigation or audit requirements. Where there may be data users not known to the Data Steward and/or Data Manager, a notice of intent to destroy the data and the proposed destruction date should be published through the relevant data access points. The notice should provide a contact for users who have reason to believe the data should be retained. Any requests to maintain the data must be resolved prior to destruction. Where possible, an automated destruction process through data tagging should be implemented to ensure destruction occurs as scheduled.

[i] https://cybersecurity.yale.edu/mss/7/7
[ii] https://gdpr-info.eu/issues/privacy-by-design/

[iii] https://cybersecurity.yale.edu/spa
[iv] https://cybersecurity.yale.edu/mss
[v] https://cybersecurity.yale.edu/support/request-exception