PennState
Institute for CyberScience

ICS-ACI Policy Series

ICS-ACI-P020 Data Protection and Retention

This is part of a series of documents that make up the formal policies adopted by the Institute for CyberScience at the Pennsylvania State University.

Last Updated: January 26, 2017

Version History:

Date	Version	Name	Description
	1.0		Initial Release
1/26/17	1.1	Avery Urusow	Formatted document with ICS theme.

Contents

1.0 Overview

The purpose of this document is not to impose restrictions that are contrary to the Pennsylvania State University's established culture of openness, trust, and integrity. The Institute for CyberScience Advanced Cyber Infrastructure (ICS-ACI) is committed to protecting our employees, partners, and the University from illegal or damaging actions by individuals, either knowingly or unknowingly.

Ultimately, it is data that is the property of the Pennsylvania State University that must be protected no matter where it resides. Effective security is a team effort involving the participation and support of every University employee and affiliate who deals with data and/or data processing systems. It is the responsibility of every resource user to know these requirements and guidelines and to conduct their activities accordingly.

2.0 Purpose

The purpose of this policy is to outline the protection of data that is created, collected, or manipulated by personnel that fall within the scope of ICS-ACI at the Pennsylvania State University. Improper handling of data exposes ICS-ACI and the University to risks including virus and malware attacks, compromise of network systems and services, data loss or contamination, and numerous potential legal issues.

3.0 Scope

This policy applies to any person who utilizes resources that are managed by ICS-ACI and to any person who handles data that is managed by ICS-ACI. The policy augments the following Pennsylvania State University policies, as well as additional applicable policies:

- AD11 – University Policy on Confidentiality of Student Records
- AD20 – Computer and Network Security
- AD23 – Use of Institutional Data
- AD35 – University Archives and Records Management
- AD71 – Data Categorization
- ADG01 – Glossary of Computer Data and System Terminology
- ADG02 – Computer Facility Security
- ADG07 – Data Categorization Examples
- General Retention Schedule (Formerly Appendix 18)
- Government Policies – HIPPA and FERPA

4.0 Policy

4.1 Philosophy

University policies regarding data protection and retention as specified here are not exhaustive, but rather serve as a foundation that resource administrators and data stewards will build upon to ensure that research information is adequately protected. Some research sponsors have their own requirements for data protection and retention that may exceed the foundational requirements provided for by the University. In these instances, the grantee (PI) must ensure that the sponsor's requirements are being met and that the associated higher-level protection measures are in place and fully operational. ICS-ACI consultants can work with PIs to help them accomplish these tasks. In these cases, the PI is expected to be aware of the requirements of the funding agency or research sponsor and to provide the ICS-ACI consultant with the necessary information and contacts that will allow them to provide the best possible assistance.

4.2 Responsibilities

It is the responsibility of the University to ensure that research is conducted ethically and that data associated with the research is properly maintained and is readily available for evaluation and further investigation.

The Principal Investigator (PI) is the steward of the data associated with their research and is responsible for ensuring that the conditions of the grant under which the research is being conducted are being met. The PI is ultimately responsible for retaining research data for a period that meets both University requirements and the requirements of any grant sponsor(s).

ICS-ACI is the steward of the University-provided ICS resources that the PI and their team are using as they conduct their research. ICS-ACI has a defined level of service that they provide when a PI signs on to use ICS-ACI services. ICS-ACI may not be able to meet every possible PI request or need. ICS-ACI will make a best effort to help PIs above and beyond the defined level of service, consistent with a fair allocation of resources and staff effort. PIs will need to provide adequate information for ICS-ACI to determine whether or not out-of-scope services can be provided, and if so, at what cost.

ICS-ACI is not responsible for third-party software or web applications that are developed and deployed by non-ICS-ACI staff. Any third-party software that is hosted on ICS-ACI-owned assets or University-owned systems should not create a security risk. Any third-party service that is hosted on ICS-ACI-owned assets or University-owned systems must meet the minimum University security requirements. We encourage you to use Penn State Security Operations and Services (SOS) resources to ensure that your services are properly designed, developed, and deployed, and that associated data meets minimum protections based upon its categorization.

It is everyone's responsibility to report suspected data protection violations to ICS-ACI staff. If you report violations to ICS-ACI staff, you are encouraged to follow up to ensure that any necessary upstream reporting to Penn State SOS and mitigation has occurred.

4.3 Data Categorization

Penn State requires an adherence to Policy AD71 Data Categorization by anyone who handles data in any way. AD71 is further used as the foundation for the University's Minimum Security Baseline, which outlines how systems and networks must be architected in order to protect the different levels of data identified in AD71. The University has also published a guideline, ADG07, which provides examples of Data Categorization.

4.4 Software Installations

No initial software installations should occur on ICS-ACI-owned systems without first coordinating with ICS-ACI staff. All software distribution sources must be verified as safe and secure before they may be used to download and/or install software. All software must be used under the terms outlined in the product's licensing documentation and/or terms of use agreement.

4.5 Data Retention

AD35 University Archives and Records Management references the General Retention Schedule (previously Appendix 18), which in turn contains a table specific to "Grant and Contract Records." This table lists "Scientific and Technical Data" as having a retention policy of "3 years in Office; Transfer to Archives for PERMANENT file." ICS-ACI does not have the resources to provide permanent retention for all research data. The data retention information in this section is specific only to ICS-ACI storage resources provided as part of our operating environment. PIs are responsible for ensuring that their research data that is resident on ICS-ACI resources is ultimately stored on other media that meets the requirements of the University and their research sponsor(s).

ICS-ACI provides the following common storage locations for system users. The table below identifies each of these locations and their respective data retention periods.

Storage Location	Retention
Home Directory	One year after termination of SLA or one year after account termination, whichever is longer.
Work Directory	One year after termination of SLA or one year after PI's account termination, whichever is longer.
Group Storage	One year after termination of SLA or one year after PI's account termination, whichever is longer.
Scratch Directory	Max 30 days; may be purged in exceptional circumstances.

4.5.1 Resources

PIs have several resources available at the University to discuss options for meeting data retention requirements, including:

- The ICS-ACI i-ASK Service Center (iask@ics.psu.edu)
- ICS-ACI consultants
- Penn State University Libraries
- Other PIs that have been subject to similar requirements

4.5.2 Data Backups

Files in the Home, Work, and Group filesystems are backed up once every 24 hours; these backups are available for recovery for 90 days. Recovery of data that is backed up requires assistance from ICS-ACI staff and the submission of a support request through the ICS-ACI i-Ask Service Center.

4.5.3 Scratch Data

Data in "Scratch" locations is NOT BACKED UP and cannot be recovered if purged. This data is automatically purged every 30 days, but can also be purged within the 30-day period due to maintenance actions or other special circumstances.

4.5.4 Transfer

PIs that are transferring outside of the University may request that a duplicate set of their data be transferred with them. ICS-ACI will work with PIs regarding the mechanism for transfer. Note that additional costs may apply (e.g., if data is provided on physical hard drives).

ICS-ACI personnel cannot transfer data that resides in a Home folder to anyone other than the owner of that Home folder. If a request is received for data from a Home folder that is not owned by the requestor, ICS-ACI personnel will provide a copy of the data to Penn State Security Operations and Services (SOS), who will sanitize the data and work with the requestor to satisfy their request.

The Home folder is not for sharing; it is a container for the personal files and application data of the folder owner. Sharing should occur out of Work and Group folders and Scratch space as required. Users will not change permissions on their Home folders so as to allow access by other users. Note that some ICS-ACI system administrators have permission to access any data on ICS-ACI storage devices. ICS-ACI system administrators will only access other users' data when it is absolutely required as part of their job functions. ICS-ACI personnel are subject to the same sanctions as any other user who misuses ICS-ACI resources.

4.6 Compromise Response

If you suspect that a data compromise may have occurred, or if you identify a situation that could potentially lead to a data compromise, then it is your responsibility to report it. The following reporting structure should be followed with progression down the list if you feel that the entity you are reporting to is not responding adequately to the situation:

- Principal Investigator
- ICS-ACI Security
- ICS-ACI Administrative Staff
- Penn State Security Operations and Services

When ICS-ACI is made aware of a potential compromise, any potentially compromised account is labeled “Inactive” and any potentially compromised node is removed from network connectivity and seized. If Penn State Security Operations and Services is not already aware of the compromise, then a report is submitted to them as well. If the compromise is validated, then the node is fully scanned for any Personally Identifiable Information (PII) or other critical data, and a report is submitted to SOS. Once mitigation instructions have been provided by SOS, the node is wiped and rebuilt.

5.0 Enforcement

Any employee, student, or visitor found to have violated this policy may be subject to disciplinary action by their Administrative unit, the College, or the University.

6.0 Supporting Documents

ICS-ACI-P000: Introduction to ICS-ACI Services

ICS-ACI-P001: Acceptable Use

ICS-ACI-P030: Authentication and Access Control

7.0 GLOSSARY

ACI-b	ICS-ACI sub-system configured to execute jobs submitted to a variety of queues, i.e. batch processing.
ACI-u	ICS-ACI User-Specific “Development/Test” interactive subsystem where PIs may specify a system configuration for user-specific interactive sessions, including root access and user-defined software stack.
Batch	Executing or processing of a series of programs (jobs) on a system without manual intervention.
Core	Data processing unit within a server. The total cores per server is dependent upon the vendor’s architecture of the server.
Core Allocation	Amount of physical compute resources purchased by or granted to a user through ICS-ACI plans.
F&A	Facilities and Administration charge, sometimes referred to as “indirect” or “overhead”.
GPFS	General Parallel File System.
Group	A self-defined set of multiple users—for example, students and researchers in a faculty member’s lab. Such rights as access to storage and allocation of resources can be delegated in an organized fashion by the PI.
Group Storage	Dedicated disk space for storing group-related data or research.
Guaranteed Response Time	The maximum time that it takes for a job to start execution after submission to a queue.
Home Directory	A user’s dedicated disk space for storing personal files, directories and programs. Directory that a user is taken to after logging into the system.
ICS-ACI	Institute for CyberScience - Advanced Cyber Infrastructure.
ICS-ACI-Burst	Queue to allow usage of compute resources in the ACI-b subsystem above a PI’s physical allocation that are needed for a short time period.
ICS-ACI-Guaranteed	Queue providing access to the ACI-b subsystem within a guaranteed time, provided request is within a PI’s physical allocation.
ICS-ACI-Open	Queue to provide user access to idle ACI-b resources that can be used during times when supply exceeds demand.
Legacy Systems	Pre-2015 ICS computing systems, such as the Lion-X clusters.
Login Nodes	Front-end servers used to log in to the ICS-ACI compute system.
NAS	Network-Attached Storage.

PI	Principal or Primary Investigator. Person, such as faculty, who is authorized to direct all of his or her research ICS-ACI resources, e.g. access, storage, compute.
Pre-emption	The act of pausing or stopping a job that is currently processing in order to fulfill terms and conditions to other users under service level agreements.
Scratch Directory	Disk space dedicated for temporary storage of data.
Service Level Agreement (SLA)	Agreement between ICS and Research PI in relation to research ICS-ACI resources, e.g. access, storage, compute.
Subsystem	A unit or device that is part of a larger system, e.g., ACI-b.
System	The computing engine along with the software, storage, network, and peripheral devices that are necessary to make the computer function, e.g., ICS-ACI.
User	A person, such as a student or faculty, who has a user account to use the ICS-ACI resources.
User Account	The means by which a user can access a computer system. ICS-ACI has four distinct user accounts: PI, Student, Staff, and Sponsored Guests.
Wall Time	A queue parameter that is set to define the maximum allowable execution time for a job once it has started.
Work Directory	User's dedicated disk space for storing research data.

PennState
Institute for CyberScience

203 Computer Building

The Pennsylvania State University

University Park, Pa 16802

Email: ics@psu.edu

Website: https://ics.psu.edu