Data Classification Policy

References

- Effective Date: June 6, 2025
- Approved By: P20 WIN Data Governing Board

Policy Purpose

The purpose of this policy is to define fundamental concepts for P20 WIN data classification and provide language that can be used by all participants to describe the ways in which data are classified and protected. As defined in the National Institute of Standards and Technology IR 8496 Data Classification Concepts and Considerations for Improving Data Protection; *Data classification is the process an organization uses to characterize its data assets using persistent labels so those assets can be managed properly [1] (NIST IR 8496, 2023).*

Classifying data is key to improving security and confidentiality for P20 WIN Data Requests. In the event of an incident, data classifications assist in determining the level of risk associated with exposure of data. It also helps to establish guidelines for applying the right security measures and controls for data access and movement within the Secure Data Enclave.

Definitions

The definitions below are meant to serve as universal key terms for describing the data classification process that is specific to P20 WIN.

Data Classification: The process an organization uses to characterize its data assets using persistent labels so those assets can be managed properly.

Data Classification Policy: An organization's data classification scheme and the formal description of the data types within the organization.

Data Classification Scheme: A taxonomy of all of an organization's known data asset types.

Data Element: A basic unit of information that has a unique meaning and subcategories of distinct value.

Data Set: Any collection of data created for the purposes of a data request.

Public Data: Data that is made freely available to the public and can be openly accessed by anyone.

Sensitive Data: Data that has the potential for harm if improperly accessed or disclosed.

Background

The P20 WIN longitudinal data system is the mechanism by which data from multiple agencies are matched to address critical policy questions. P20 WIN is used to inform sound policies and practice through the secure sharing of longitudinal data across Participating Agencies to ensure that residents successfully navigate supportive services and educational pathways into the workforce.

Connecticut General Statutes 10a-57g codifies P20 WIN and "its purpose to establish processes and structures governing the secure sharing of critical longitudinal data across Participating Agencies through implementation of the standards and policies of the P20 WIN network." This statute also establishes a data governance structure "to establish and enforce policies related to cross-agency data management, including, but not limited to, data confidentiality and security in alignment with the vision for CP20 WIN and any applicable law. In establishing such policies, the data governing board shall consult with the Office of Policy and Management, in accordance with the provisions of section 4-67n and other applicable statutes and policies."

All Participating Agencies are responsible for classifying the data sets they collect and provide for the purposes of fulfilling P20 WIN Data Requests.

Scope of Policy

This policy shall apply to data sets used for approved P20 WIN Data Requests as determined by P20 WIN Participating Agencies, P20 WIN Operating Group, the Data Integration Hub staff, and any other entity that provides, shares, and accesses data for an approved data request.

Data sets shared through P20 WIN is governed by several legal, regulatory, and compliance obligations some of which require that all sensitive data is classified as a measure of protection.

The obligations outlined below are not intended to be exhaustive and are subject to revision in response to evolving federal, state, agency, and industry regulations.

Type of Data	Relevant federal and state laws	Agency Coverage
Child Welfare	Title IV-E and IV-B, CAPTA, HIPAA, IDEA Part C	DCF, OEC
Criminal Justice	CT Gen Stat (C.G.S) § 46b-124, C.G.S. §§ 54-142m, C.G.S. §§ 51-36a, C.G.S. §§ 17a-28, HIPAA, FERPA, 42 U.S.C. § 290dd-2, CJISD-ITS-DOC-08140-5.8	JBCSSD, DOC
Financial Aid	Higher Education Act (FAFSA and student loans), Internal Revenue Code	OHE, CCIC, CSCU, UCONN
Education	FERPA	OEC, CSDE, OHE, CSCU, CCIC, UCONN, DCF
Health (including mental health)	HIPAA, C.G.S. § 52-146e-j	DSS, DMHAS, DCF, DOC, OHE, CCIC, CSCU, UCONN, OEC
Homelessness	HUD HMIS regulations and standards	CCEH (DOH / DMHAS)
Medicaid	Medicaid, HIPAA, C.G.S. § 17b-90	DSS
SNAP	SNAP federal law (7 C.F.R. § 272.8(a)(4))	DSS
Substance Use	42 C.F.R. Part 2	DMHAS, DSS, DCF
TANF	TANF federal law (45 C.F.R. § 205.50)	DSS
Workforce and wages	Unemployment compensation (20 C.F.R. 603), C.G.S. § 31-254, WIOA, TAA, JFES, Apprenticeship	DOL

Classification Scheme

The classification scheme adopted for P20 WIN is based on the sensitivity level associated with each data set created at varying stages of a P20 WIN Data Request, and the risk severity in the event of data exposure. The P20 WIN classification scheme defines four levels of data classification. The classification of data sets under this policy will be informed by a standardized set of guidelines that define when, how, and by whom data classification is conducted. These guidelines will be documented in the P20 WIN Data Classification Methodology to be maintained and updated by the P20 WIN Operating Group in collaboration with the P20 WIN Participating Agencies.

Classification Levels

Public: Data is **openly available** and appropriate for public consumption. Data in this classification poses **no risk** to an individual or the P20 WIN longitudinal data system if exposed.

Internal: Data is **internal to the P20 WIN longitudinal data system** and access is restricted to Participating Agencies, the P20 WIN Operating Group, and approved data requesters only. Data in this classification poses **no risk to an individual** and **minimal risk** to the P20 WIN longitudinal data system if exposed. **Minimal** reputational, financial, operational, or legal harm is anticipated for P20 WIN in the event of unauthorized disclosure.

Confidential: Data is **internal to the P20 WIN longitudinal data system** and access is restricted to those whose need to know is relevant to their operational responsibilities or decision-making authority within the P20 WIN longitudinal data system. Data in this classification poses a **moderate risk** to an individual or the system if exposed. **Some** reputational, financial, operational, legal, psychological, or discriminatory harm is anticipated for an individual or the P20 WIN longitudinal data system in the event of unauthorized disclosure.

Restricted: Data is **personal to an individual or internal to the P20 WIN longitudinal data system** and access is highly restricted to those whose need to know is relevant to their decision-making authority within the P20 WIN longitudinal data system. Data in this classification poses a **significant level of risk** to an individual or the P20 WIN longitudinal data system if exposed**. Considerable and irreversible** reputational, financial, operational, or legal, psychological, or discriminatory harm is anticipated for an individual or the P20 WIN longitudinal data system in the event of unauthorized disclosure.

Considerations for Classifying Data

Participating Agencies are responsible for classifying the data sets their agency provides into one of the P20 WIN data classification levels. It is important to note that datasets go through various changes throughout the Data Request process, thus impacting their associated risk and classification level. The P20 WIN Data Classification Methodology document will address the changing nature of data throughout the P20 WIN data request process.

To determine which level a particular dataset falls into, agencies should consult the P20 WIN Classification Methodology and consider the following characteristics of the data, as recommended by NIST SP 800-122, Section 3.2.

1. *Identifiability*
    a. Agencies should consider how easily a particular data element can be linked to an individual.
    b. Agencies should consider which combination of data elements increase the ease of identifying an individual.
2. *Quantity of PII*
    a. Agencies should consider whether the risk associated with larger volumes of PII warrant a stricter classification level.
3. *Data Element Sensitivity*
    a. Agencies should consider the sensitivity of certain data elements particularly when combined with other data elements.
4. *Context of Use*
    For classifying data, context of use is defined as the purpose for which PII is collected, stored, used, processed, disclosed, or disseminated.
    a. Agencies should consider how the context of use for particular data elements may subject an individual to potential harm.
    b. Agencies should also consider whether the disclosure of collection and use of PII on its own may subject an organization or individual to potential harm.
5. *Obligations to Protect Confidentiality*
    a. Agencies should consider the laws, regulations, and internal policies they are subject to with respect to protecting confidentiality.
6. *Access to and Location of PII*
    a. Agencies should consider the levels of access to PII within their organization and where those PII are located. Organizations may restrict access to PII, and minimize where it is stored, as measures to reduce the risk associated with it.
7. *Level of Harm*
    a. Agencies should consider the level and type of potential harm to an individual in the event of data exposure.
    b. Agencies should consider the level and type of potential harm to the system or their agency in the event of data exposure.

While Participating Agencies and external data requesters may have their own data classification standard and policies, it is the expectation that the classification levels applied to and referenced throughout the P20 WIN Data Request process, are in alignment with the levels specified in this policy. Participating Agencies must collaborate to consider all of the recommended characteristics above and determine if classification levels should be enhanced for any data set at the start of a data request. Data classifications cannot be lowered below the minimum default classification for each data set.

Related Documents

The following document list references other data policies, resources, and the related process documents:

Connecticut Data Privacy Breach Law

CSCU Data Classification Policy

CTECS Data Classification Policy

UConn Data Classification Policy

Revision History

Version Number	Version Date	Description of Change	Point of Contact
1	6/6/2025	Adopted by Data Governing Board	Katie Breslin

Sources

1. NIST IR 849: Data Classification Concepts and Considerations for Improving Data Protection

2. NIST Special Publication 800-53: Security and Privacy Controls for Information Systems and Organizations

3. NIST SP 800-171 Rev. 3: Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations

4. Daniel J. Solove, *Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 Nw. U. L. Rev. 1081 (2024). https://scholarlycommons.law.northwestern.edu/nulr/vol118/iss4/4