

# Rhode Island Longitudinal Data System

## Data Quality & Metadata Policies

*Draft May 28, 2024*

### 1. Purpose

Rhode Island Longitudinal Data System (“RILDS”) Data Quality & Metadata Policies are articulated in the Rhode Island Longitudinal Data System Act (“RILDS Act”) as well as each of the data sharing agreements that have been established between the Rhode Island Longitudinal Data System Center (“Center”) and the Data Contributing Agencies. These policies establish how data integrated and linked through RILDS will be managed so that data are accurate, complete, consistent, updated regularly, valid, and accompanied by appropriate metadata.

The RILDS Center operates and maintains RILDS in accordance with federal and state laws and regulations particularly regarding education, health, and employment records, including, but not limited to the Family Educational Rights and Privacy Act (FERPA), the Individuals with Disabilities Education Act (IDEA), the Health Insurance Portability and Accountability Act (HIPAA). The following definitions pertinent to addressing data quality are derived from these and other related documents.

### 2. Definitions

#### 2.1. Data Contributing Agencies

Data Contributing Agencies are defined as any Rhode Island state agency, Rhode Island municipal government, research institution, non-profit organization, community foundation, or other entity that has executed the RILDS memorandum of understanding (MOU).

#### 2.2. Data Governance Committee

The RILDS Data Governance Committee established per R.I. Gen. Laws § 42-165-5(b)(6) with authority to approve Requests.

## 2.3. Executive Governing Committee

The RILDS Executive Governing Committee established per R.I. Gen. Laws § 42-165-5 with responsibility for governing RILDS.

## 2.4. Non-Data Contributing Agency

Any Rhode Island State agency, excluding Data Contributing Agencies, that serves on the Executive Governing Committee as named in the Rhode Island Longitudinal Data System Act, namely the Rhode Island Department of Administration, Rhode Island Division of the Information Technology, Rhode Island Executive Office of Health and Human Services, and Rhode Island Office of Management and Budget.

## 2.5. Personally Identifiable Information

Information that alone is linked or linkable to a specific individual and that would allow a reasonable person to identify the individual with reasonable certainty.

## 2.6. Public Output

Any presentation, publication, report and other product ready for public release containing aggregate-level RILDS data, such that individuals cannot be directly or indirectly identified. In other words, Public Outputs do not contain Personally Identifiable Information.

## 2.7. RILDS

The data system formerly known as the Rhode Island DataHUB operated by DataSpark at the University of Rhode Island, is the State of Rhode Island's statewide longitudinal data system, established per R.I. Gen. Law § 42-165-2.

## 2.8. RILDS Center or Center

The entity formerly known as DataSpark at the University of Rhode Island, and whatever other resources are necessary to accomplish the powers and duties prescribed in R.I. Gen. Law § 42-165, manages and operates RILDS and conducts research and evaluates programs.

## 2.9. RILDS MOU

RILDS MOU refers to the memorandum of understanding or other agreement executed by and between the RILDS Center and any Rhode Island state agency, Rhode Island municipal government, research institution, non-profit organization, community foundation, or other entity for recurring participation in RILDS.

## 2.10. State and Federal Privacy Laws

All applicable state and federal laws and accompanying regulations, including but not limited to the federal Family Educational Rights and Privacy Act and its accompanying regulations (“FERPA”), Health Insurance Portability and Accountability Act (“HIPAA”), R.I. Gen. Laws § 28-42-38, 20 CFR 603.1 et seq., and any other privacy measures that apply to the personally identifiable information that is used by the RILDS Center and/or becomes part of the RILDS.

# 3. Data Collection & Preparation

## 3.1. Policy Statement

Per the RILDS Data Governance Policies, the Data Contributing Agencies’ Data Stewards are responsible for updating records. In the update process, Data Stewards will prepare records with the least manipulation possible and the most fidelity possible to Data Contributing Agencies’ source data systems. The Data Stewards will review and, as necessary, revise associated metadata and processing documentation.

## 3.2. Scope

The above policy governs all data regularly integrated into RILDS.

## 3.3. Reason for Policy

Regular updates ensure data is available when needed. Preparing data with the least manipulation possible and the most fidelity possible to source systems increases state efficiency and improves data timeliness. Data Stewards can more quickly prepare extracts, and the RILDS Center can more quickly integrate and link these records.

Documenting necessary processing ensures knowledge is not only transferred between the Data Contributing Agencies and the RILDS Center but also retained when Data Stewards leave state employment. It also contributes to the development of a statewide culture of data quality and Data Contributing Agencies’ internal data governance.

## 3.4. Processes

The Data Contributing Agencies’ representative to the Data Governance Committee names at least one (1) Data Steward for each data collection shared through RILDS. This Data Steward may be the Data Contributing Agencies’ representative to the Data Governance Committee, as long as the Data Steward can fulfill the responsibilities outlined in the RILDS Data Governance Policies.

The Data Steward updates records, providing new data extracts, at least annually but not more frequently than monthly. The Data Steward communicates this schedule, including both frequency and deadlines, to the RILDS Center.

The Data Steward provides metadata for each data collection shared through RILDS as a .csv or .xlsx file. At a minimum, these metadata include the elements' names, exactly as the elements appear in the extract, and the elements' definitions. These metadata may also include the elements' data types, values, and those values' definitions, as well as identify primary and foreign keys.

Ten (10) business days before collection submission deadlines, the Center reminds the Data Steward via email, copying the Data Contributing Agency's representative to the Data Governance Committee, that an update is required. The Center provides the file layout, any processing documentation previously shared by the Data Steward, and the associated metadata.

In the update process, Data Steward prepares records with the least manipulation possible and the most fidelity possible to Data Contributing Agency's source data systems. The Data Steward documents any necessary processing, identifying source elements and describing any calculations, deduplications, formatting changes, joins, and other processing, and shares this documentation with the RILDS Center. In the update process, the Data Steward reviews and, as necessary, revises these metadata and this processing documentation. The Data Steward submits extracts according to the processes outlined in the RILDS MOU.

When the file layout has changed, the Data Steward and the Center follow the processes detailed in the Data Modification Policy.

## 4. Data Quality

### 4.1. Policy Statement

Upon receipt of data extracts from Data Contributing Agencies, the RILDS Center performs data quality checks. The Center proceeds with integrating and linking these extracts, unless the Center discovers that the extract contains elements not identified in the RILDS MOU.

### 4.2. Scope

The above policy governs all data regularly integrated into RILDS and data integrated temporarily for project-specific use.

### 4.3. Reason for Policy

By confirming that only elements identified in the RILDS MOU are integrated and linked through RILDS, these checks support compliance with the RILDS Privacy & Security Policies. The checks also ensure that data are accurate, complete, consistent, and valid.

The Center ensures data availability by integrating and linking extracts with only minor data quality concerns.

## 4.4. Processes

Upon receipt of data extracts from Data Contributing Agencies, the RILDS Center performs data quality checks, ensuring that:

- no unexpected data elements are provided;
- all expected data elements are provided; and
- record or row counts are consistent with previous extracts.

The Center completes these checks within fifteen (15) business days of receipt.

When the Center discovers unexpected data elements are provided, the Center immediately destroys the extract, retaining only the names of the unexpected elements. The Center notifies the Data Steward, copying the Data Contributing Agency's representative to the Data Governance Committee. This notification identifies the unexpected elements. The Center asks the Data Steward whether the file layout has changed. When the Data Steward confirms that the file layout has changed, the Data Steward and the Center follow the processes detailed in the Data Modification Policy. When the file layout has not changed, the Center requests a corrected extract. This corrected extract contains only elements identified in the RILDS MOU. The Data Steward is responsible for transferring a corrected extract within fifteen (15) business days of notification.

When the Center discovers that expected data elements are not provided, the Center notifies the Data Steward, copying the Data Contributing Agency's representative to the Data Governance Committee. This notification identifies the missing elements. The Center asks the Data Steward whether the file layout has changed. When the Data Steward confirms that the file layout has changed, the Data Steward and the Center follow the processes detailed in the Data Modification Policy. When the file layout has not changed, the Center proceeds with import and requests a corrected extract. This corrected extract contains all elements in the data collection. The Data Steward is responsible for transferring a corrected extract within fifteen (15) business days of notification.

When the Center discovers the record or row counts are inconsistent with previous extracts, the Center asks the Data Steward to confirm whether the extract is complete (e.g., were records mistakenly excluded?). Record or row counts are inconsistent when the counts differ by more than one standard deviation from the mean number of records in previously submitted extracts covering a similar period. When the Data Steward confirms that the extract is complete, the Center asks the Data Steward to document the programmatic or policy changes that affected the record or row counts (e.g., middle schools are now required to report grades earned where previously only high schools reported grades earned). When the Data Steward confirms that the extract is missing records, the Center proceeds with import and requests a corrected extract. This corrected extract contains all records for the period. The Data Steward is responsible for transferring a corrected extract within fifteen (15) business days of notification.

## 5. Data Modification

### 5.1. Policy Statement

Upon the Data Steward's confirmation that the file layout has changed, the RILDS Center modifies the Data Contributing Agency's data shared through RILDS according to the terms specified in the RILDS MOU.

### 5.2. Scope

The above policy governs all data regularly integrated into RILDS.

### 5.3. Reason for Policy

By confirming that only elements identified in the RILDS MOU are integrated and linked through RILDS, these checks support compliance with the RILDS Privacy & Security Policies.

### 5.4. Processes

Upon the Data Steward's confirmation that the file layout has changed, the RILDS Center modifies the Data Contributing Agency's data elements shared through RILDS. This modification proceeds according to the terms specified in the RILDS MOU.

In most instances, the Center prepares a memorandum, notifying the Data Contributing Agency's Point of Contact named in the RILDS MOU. Where the Point of Contact is not the Data Contributing Agency's representative to the Data Governance Committee, the Center copies that representative. This memorandum identifies elements that have either been removed or added. The Center attaches the full list of the Data Contributing Agency's data elements shared through RILDS.

The Data Contributing Agency's Point of Contact signs and returns this memorandum, authorizing modification. The Center retains all records and elements, until or unless the Data Contributing Agency requests destruction.