CASE WESTERN RESERVE UNIVERSITY

Enterprise Data Standards and Duplicate Record Prevention Guidelines

Faculty Information System

Human Capital Management System

Student Information System

July 2023

Executive Summary

With the addition of the Faculty Information System (FIS), there are now three systems sharing basic data elements for students, staff, faculty, and other university constituents: the Student Information System (SIS), the Human Capital Management System (HCM), and FIS. There is an interface process that transfers data among the systems and to the identity management system (IDM) so that university NetIDs can be assigned and access to other university resources can be properly managed.

Maintaining accurate data is critical to all users of university systems. Accurate data is essential for making informed decisions, maintaining legal and regulatory compliance, and building trust in our data sources as credible and valid. However, and most importantly for the purposes of this document, accurate data enables us to provide the best levels of service to all of our constituents.

This document provides

- Background and context for the need for guidelines
- Basis for understanding why these standards are important
- Description of critical data elements for each of the three enterprise systems
- Framework for establishing stakeholder accountability

There are a variety of entry points for data into each system, whether manual or by interface. In each system, there are several access roles with permission to create new records. This document is intended to provide standards to which all users with records creation access must adhere.

Several people representing functional ownership for each of the three systems met to establish a mutual understanding of each system and to agree to the standards and processes described herein. Representatives from other stakeholder groups were consulted and have also agreed to these standards and processes (see Appendix A).

CASE WESTERN RESERVE UNIVERSITY

Background

Systems and data sharing

Since 2008, the Oracle PeopleSoft Student Information System (SIS) and the Human Capital Management System (HCM) have shared basic data elements across systems via an interface. As a result, people with multiple roles (e.g., student employees, staff/faculty enrolled in classes, etc.) have a single EmplID (the seven-digit unique identifier used by SIS and HCM) assigned, and their information is maintained in one system to share with the other.

In 2023, a third enterprise-level system was added to the interface process: the Faculty Information System (FIS). When the SIS/HCM interface went live, it was discovered that records with minimal/ambiguous data were entered in HCM. Since the HCM system's go-live pre-dated SIS, there had never been any discussion of university-wide data standards. As a result, thousands of records exist across the systems with only a small amount of data (e.g., name only plus a fake birthdate and fake SSN).

What is a duplicate record? What is a merged record?

There are two types of issues we seek to avoid: duplicate records and merged records. A **duplicate record** is a record that is created for a person that in some way already exists in one of the three systems. A **merged record** is the inadvertent merging of data from one distinct individual onto the record of another, different, distinct individual.

Duplicate records

When a person has a new role with the university and data needs to be updated, it is often impossible to know if the existing record is for the same person. Therefore, it is necessary to create a new record, which may be a duplicate record. Duplicate records can and do result in:

- real data isolated/stranded with ambiguously defined record(s)
- two EmplIDs for the same person
- two network IDs (the abc123 unique identifier used by the identity and access management system) for the same person
- any combination of the above

So far, more than 1000 sets of duplicate records have been identified and corrected and there are thousands more that cannot be corrected due to insufficient identifying information. It takes from one hour to several days to address each duplicate record. Some can be resolved by a single person, while others require many people across several university departments to resolve.

Merged records

Perhaps even more problematic, ambiguous records can result in incorrectly merged data, which could violate the privacy of either party, enabling one person to see the data of another, tax data reported for the wrong person, grade information released to the incorrect individual, etc. This is a significant risk to the university and must be avoided.

CASE WESTERN RESERVE UNIVERSITY

Business Impacts

Why is it important to avoid duplicate and merged records?

Customer Impacts are Substantial
Duplicate or merged records across FIS/HCM/SIS can have disastrous customer service impacts and can also result in duplicate NetIDs.

Merged record examples:

- one person receives a transcript with a combination of their own and someone else's grades
- one person able to view another person's personal data including birth date and Social Security number (SSN)
- one person able to view another person's payroll and tax information

Duplicate record examples:

- someone may receive only a partial transcript
- only partial teaching and advising data reported to FIS; inaccurate Faculty Activity Report
- paid under two different SSNs
- student financial data may exist under two different IDs; credits under one ID, past due balance under another

Multiple Resources Required for Repairs
Correcting a duplicate record in SIS requires substantial resources, regardless of origin. In the best-case scenario, identifying and correcting a duplicate record on the same day takes about 30 uninterrupted minutes by one University Registrar staff member. It necessitates in-depth knowledge of SIS data, table structure, relationships, and business processes. In the worst case, it can take over 45 person-hours and involve multiple system admins, leading to processing delays if resources are unavailable. Systems and departments involved include SIS, HCM, FIS, IDM, UTech, PowerFaids, Terra Dotta/SEVIS, Access Services, Financials, Grants, Canvas, etc. Some systems may never receive updates due to static data batches or missed opportunities.

Opportunity Costs are High
Across all departments involved in the repair process, the users working on the records are highly skilled, highly knowledgeable, and experienced people who are relied upon by others for many other critical, priority, or specialized services which impact many other service providers. Those other services can be delayed to synchronize the manual repair of the duplicate (or triplicate) records in multiple systems.

Standards and Prevention

Representatives from the three enterprise systems met to create a set of standards to guide the collection and storage of some of the most critical pieces of information used to uniquely identify individuals across systems. In addition to addressing these inputs, the group also outlined a set of steps intended to reinforce accountability. Because entry points for data vary widely within and across systems, education and training reinforcement is a highlight of the accountability process.

Critical Data Elements

To minimize the risk of creating duplicate or merged records in the FIS, HCM, SIS, and IDM systems and to support records accuracy and good service overall, a set of critical data elements has been identified. Although all data collection and entry into our systems should be held to the highest standards, this subset of shared data elements is critical, whether records originate from the student (SIS), faculty (FIS), or staff (HCM) business processes of the university.

Minimum data elements:

- **Names** Legal name should be entered in all systems as the primary name. If the individual has a preferred name, the preferred name fields should be utilized, if available. If the individual goes by one name, use FNU or a period (.) in the first name field and the name in the last name field if the individual goes by one name. If the individual has a middle name, the complete middle name field should be utilized unless only a middle initial is available, in which case the middle initial should be entered.

- **Birthdate** End users must collect and enter an accurate, real birth date. In rare circumstances when a real date of birth is not obtainable, it is permissible to use the standard date of 01/01/1900.

- **Email address** Can be either CWRU or non-CWRU, but this is often helpful in identifying duplicate records. Use HOME email address type for non-CWRU in HCM and SIS so that it is loaded to the other systems, including FIS. Use the CASE email address type in SIS for a CWRU.

- **Mailing address** End users must enter a real home address. A home (permanent) mailing address, even if it's "CWRU Non-Employee", including city, state, and zip is required. In SIS, we can't edit a record that has a missing address, even if it's a data element that we own, e.g., campus ID or off-campus email address, unless there is a home mailing address.

- **SSN** Leave SSN blank if unknown (if available in your system). SSN isn't required in SIS, but it is requested. HCM requires SSN and must enter something in that field in

order to create an identity. When unknown in HCM, and by exception, use all 8s. When unknown in SIS, leave blank. FIS does not collect SSNs.

Note: When feasible, data collection processes should incorporate inquiries that seek additional information regarding any previous affiliations of the individual with the university, e.g., under a different name or with a different status, such as student or employee.

Framework for Establishing Prevention and Stakeholder Accountability

The best way to address the issue of duplicate records is to avoid creating them. A stepped approach has been designed to facilitate education, reinforcement, support, accountability, and remediation. Each system owner/team is responsible for user training.

Note: The performance management process should be followed to address performance concerns related to data entry. To the extent that one's job description includes data entry duties, the Second Phase/Step 3 solutions must go through proper procedures to reassign duties to someone else (revising job descriptions, evaluating salary grade designations, etc).

I. **First Phase**
   A. Introduce guidelines
      1. Provide materials to users who already have access to update relevant data in systems of record;
      2. Update process for granting new users access to include introduction of guidelines
   B. Offer training sessions
      1. Provide comprehensive training on each area's specific business process that involves the entry of new person records into their system;
      2. Facilitate involvement between data maintainers and representatives from the system of record so that process issues or system concerns related to the entry of new person records can be discussed, documented, and later addressed;
      3. Support and enhance the importance of correct and thorough business processes, and place this critical activity into proper perspective;

II. **Second Phase**
   A. Continue support through training as requested
   B. Monitor progress
   C. Address problem areas in a series of steps:
      1. **STEP 1**: when it is discovered that a user creates their first duplicate record
         a. Contact the user and discuss the specifics of the records created
         b. Offer additional one-on-one assistance or additional training (optional)
         c. Reiterate guidelines and remind the user of the next steps
      2. **STEP 2:** when it is discovered that user continues to create duplicate records
         a. Contact user and supervisor

   b. Set up additional one-on-one assistance or additional training (mandatory)
   c. Request options for a contingency plan from the department in case Step 3 is required
3. **STEP 3:** for users who have been contacted about Step 2 and continue to create duplicate records
   a. Contact supervisor
   b. Consider reassigning access to a new user to carry out functions

In some cases, it will be impossible to avoid the creation of duplicate records, given that existing records may not be complete or contain enough identifying information required to uniquely identify an individual. This should be considered when reviewing individual user situations throughout the Phase 2 process. **End users with access to create records must make every attempt to obtain accurate identifying information and should be held accountable if the information is available but not utilized.** The system owners that review each situation will seek information about each situation and review with additional team members as appropriate before recommending Step 3 of Phase 2. When new scenarios occur, system owners will update training materials as needed.

CASE WESTERN RESERVE UNIVERSITY

Appendix A

Stakeholder Involvement

Stakeholder	Review Date
UTech Identity Management Team	6/5/2023
Registrar's Committee	6/30/2023
SIS Team	7/13/2023
Data Governance Committee	7/20/2023
SIS Executive Sponsors: Don Feke and Miroslav Humer	8/14/2023
FIS Leadership Team	6/20/2023
FIS Steering Committee	7/7/2023
FIS Executive Sponsors: Eddie Bolden and Miroslav Humer	8/14/2023
HCM Executive Sponsors: Carolyn Gregory and Rose Kelly	8/31/2023
Interim Provost Joy Ward, School Deans	9/21/2023