YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

The Benefits of Collaborative IR

A Learning and Solution Seeking Workshop

Presented by the Yes We Must Coalition & Seamless Strategies Group
May 22, 2025

The Yes We Must Coalition

- 31 private, non-profit colleges and universities
- Undergraduate enrollment 50% or more Pell-eligible/dependent
- Collaborate to identify and change barriers to success for learners from poverty
- Offer workshops, discussion groups, communities of practice, facilitate communication and work among members
- Also seek grants on behalf of members
- Today's workshop is to share what 5 members are learning while implementing a grant from the Gates Foundation

YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

Enhancing PDP Adoption and Use for Under-Resourced Institutions

- $543 for 2 Years
- Ends Nov 26
- Today share the context for the grant and the first year of work and think about how we can expand the grant for further funding

YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

Audience Prompt

- If you spent less time on IPEDS/State reporting or Registrar support, what would you focus on instead?

YWMC Members Participating in the Grant

- Calumet College of Saint Joseph (HSI), Whiting, Indiana (Est. Annual Enrollment: 674)
- Coker University, Hartsville, South Carolina (Est. Annual Enrollment: 1,263)
- Heritage University (HIS on native land), Yakima, Washington (Est. Annual Enrollment: 835)
- Vaughn College (AANAPISI and HIS), Flushing, New York (Est. Annual Enrollment: 1,212)
- Villa Maria College, Buffalo, New York (Est. Annual Enrollment: 545)

History of These Members Working Together

- Context of YWMC puts focus on improving outcomes for low-income students recognizing the role of data to inform strategies/decisions
- YWM tried own data collection from members 2016-18 but didn't have the infrastructure to sustain it
- Gates grant 2019-21 to assist individual institutions to use the Postsecondary Data Partnership of National Student Clearinghouse to get the data they needed. Problems with individual submissions
- 2020 general discussions re shared positions/functions including in IR
- SeaChange grant 2022 assessment of IR capabilities and needs at the five schools that expressed interest in sharing IR

Audience Prompt

- What characteristics interest you for true peer data comparisons

Conclusions of what needed to Collaborate to Strengthen IR

- Common Data Governance Policy
- Defining Important Data Elements
- Developing Common Data Dictionary
- Automating Processes
- Common Visuals
- Common Research Questions and Peer Group Analysis

IR Processes & Our Progress

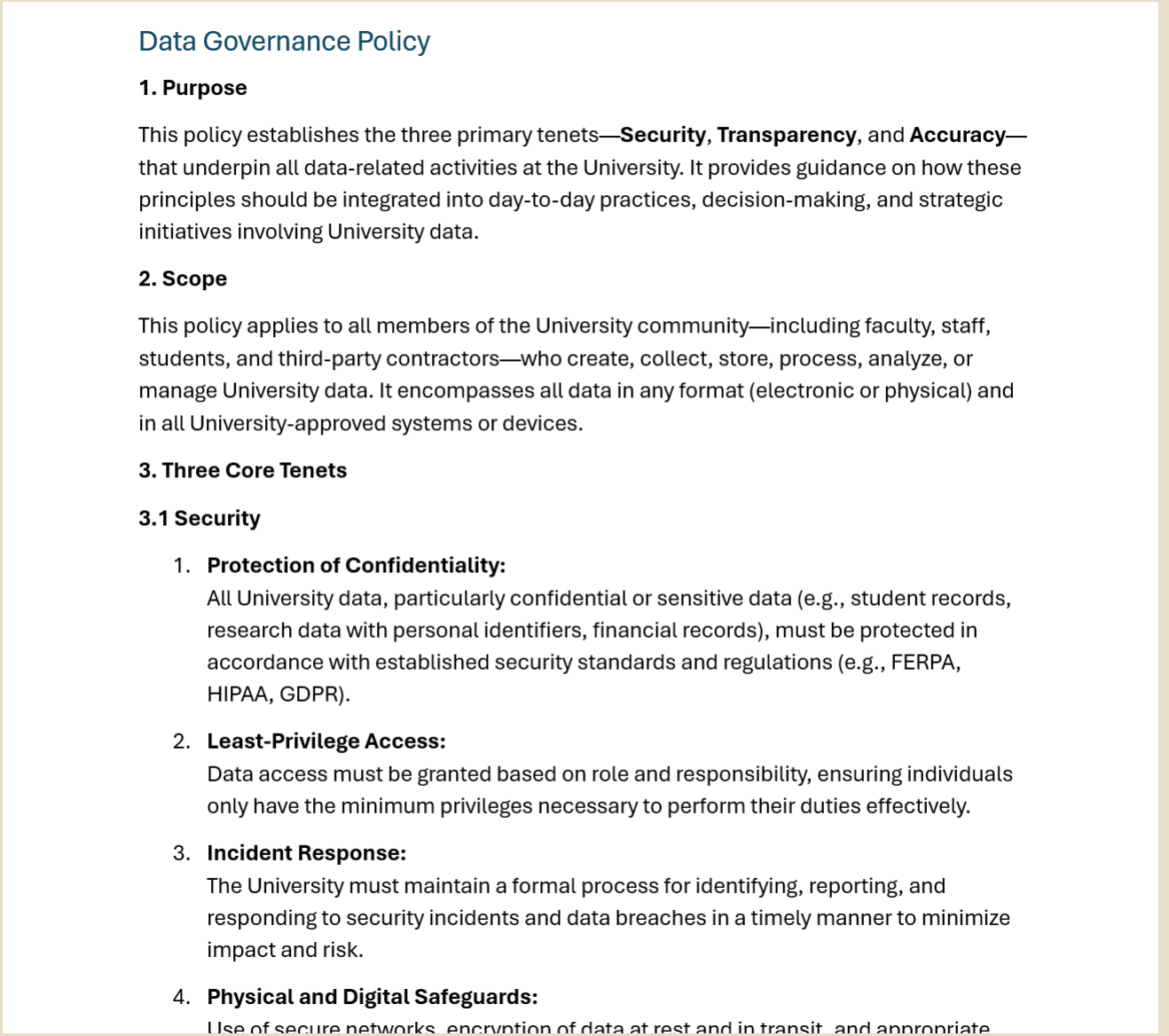
YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

IR Processes & Our Progress

- Conducted interview with IR staff and collaborators
- Learned about reporting responsibilities and data use
- Captured various reporting needs

Data Governance Overview

- Early need identified through interviews
- Drafted policy based on school needs and expert experience
- Core tenets: Security, Transparency, Accuracy
- Foundation for consistent and trustworthy



Data Governance Policy

1. Purpose

This policy establishes the three primary tenets—**Security**, **Transparency**, and **Accuracy**—that underpin all data-related activities at the University. It provides guidance on how these principles should be integrated into day-to-day practices, decision-making, and strategic initiatives involving University data.

2. Scope

This policy applies to all members of the University community—including faculty, staff, students, and third-party contractors—who create, collect, store, process, analyze, or manage University data. It encompasses all data in any format (electronic or physical) and in all University-approved systems or devices.

3. Three Core Tenets

3.1 Security

1. **Protection of Confidentiality:**
   All University data, particularly confidential or sensitive data (e.g., student records, research data with personal identifiers, financial records), must be protected in accordance with established security standards and regulations (e.g., FERPA, HIPAA, GDPR).
2. **Least-Privilege Access:**
   Data access must be granted based on role and responsibility, ensuring individuals only have the minimum privileges necessary to perform their duties effectively.
3. **Incident Response:**
   The University must maintain a formal process for identifying, reporting, and responding to security incidents and data breaches in a timely manner to minimize impact and risk.
4. **Physical and Digital Safeguards:**
   Use of secure networks, encryption of data at rest and in transit, and appropriate

YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

Data Governance In Practice – Villa Maria

- Reviewed and addressed outdated data policies
- Implemented draft governance policy with campus feedback
- Formed cross-functional assessment team (IR, faculty, admin)
- Aimed to centralize data storage and access
- Improving communication between data/report owners and IR

Common Data Dictionary

- Developed in response to data governance work
- Built using variables relevant to each school
- Shared definitions aligned across institutions
- Enables consistent benchmarking and peer comparison
- Designed to support flexible reporting and future expansion
- Foundation for report generation across schools

YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

Data Dictionary: Enrollment and Retention

seamless STRATEGIES

#	Data Element	Data Type	Description	Possible Values	Required/Optional	Notes
Student Demographic (Static)						
1	Student_ID	Alphanumeric	Unique identifier for each student	Unique alphanumeric values	Required	Used to track individual students across enrollment and retention data.
2	Date_of_Birth	Date (mm/dd/yyyy)	Student's DOB	mm/dd/yyyy	Required	Used to calculate age and age group
3	Gender	Fixed Value	Student's reported gender	Female, Male, Prefer not to specify, Unknown	Required	Useful for analyzing gender equity in enrollment, retention, and academic outcomes.
4	Race/Ethnicity	Fixed Value	Student's self-reported race/ethnicity	Nonresident Alien, American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or other Pacific Islander, White, Two or More Races, Unknown	Required	Helps institutions track diversity and identify gaps in student success.
5	First_Gen	Fixed Value	A student who is the first member of his or her immediate family to graduate from a college or university	• No parent has attended post-secondary • At least one parent has attended post-secondary but earned no credential or degree • At least one parent has a certificate • At least one parent has an associate degree • At least one parent has a bachelors degree or higher	Required	Aligns with PDP definition of first generation. Vaughn and Heritage collect boolean flag, for them a First Gen Student would be mapped to "No parent has attended post-secondary" and a Non First Gen Student to "At least one parent has attended post-secondary but earned no credential or degree"
Enrollment Data (Static)						
6	Cohort_Year	Date (YYYY)	Academic year the student first enrolled	YYYY (e.g., 2024, 2025)	Required	Helps categorize students into entering cohorts for tracking retention and outcomes.
7	Cohort Term	Fixed Value	Academic term in which the student enrolled	Fall, Winter, Spring, Summer	Required	Tracks term-level enrollment patterns.
8	Enrollment_Type	Fixed Value	The student's enrollment type at the time of registration	First-time, Transfer-in	Required	Determined by institution's credit hour thresholds.

Why PDP? A Foundational Data Set

- Selected Postsecondary Data Partnership (PDP) based on shared data priorities
- Managed by the National Student Clearinghouse
- Designed to support longitudinal, student-centered analysis
- Adopted by 700+ institutions nationwide
- Powers benchmarking and accountability efforts
- We developed tools to simplify and automate PDP file creation

Audience Prompt

- What is your experience with the PDP?

Process Improvement & Automation Opportunities

Process Improvement

- Manual processes were common across institutions
- Time-intensive reporting and data prep tasks
- Opportunities identified through interviews and hands-on support
- Upcoming solutions designed to reduce effort and improve consistency

Heritage University – Automating Recurring Reports

- No formal process for recurring reports (IPEDS, PDP, etc.)
- Manual, multi-step process: extract, clean, and reformat data
- Reports submitted manually; feedback loops often repeat the cycle
- Process is time-consuming, error-prone, and person-dependent
- YWM Automation aims to:
  - Improve efficiency and accuracy
  - Standardize recurring workflows
  - Increase transparency and sustainability

Calumet College – Streamlining Course Scheduling

- Registrar manually collects course scheduling data from faculty and program directors
- Information includes: course offerings, assigned instructors, and alignment with academic calendar
- Data is manually entered into the SIS after extensive outreach
- Current process is repetitive and time-intensive
- Proposed automation to:
  - Standardize course schedule collection
  - Streamline SIS data entry

Coker University – Streamlining IPEDS Reporting

- IPEDS requires detailed breakdowns by race, gender, and enrollment status
- Data is manually entered, time-consuming and error-prone
- Prior year reporting helps, but consistency remains a challenge
- Issues include shifting data sources, changing criteria, and staff turnover
- Python scripts based on standardized logic developed with Seamless Strategies to reduce manual entry and improve year-over-year consistency

Vaughn College – Automating Missing Grade Notifications

- SIS: Jenzabar Sonis; previously relied on system-generated emails
- IR developed an Excel report for the Registrar
- Includes course, instructor, and detailed student information
- Report updates automatically every minute during active terms
- Time intensive manual steps:
  - Missing grade reminders
  - Instructor outreach
  - Grade finalization tracking
- Opportunity identified for workflow automation

Vaughn College – Streamlining PDP Data Submissions

- Past PDP efforts stalled due to:
  - Small IR team with limited experience
  - Unfamiliarity with mapping SIS data to PDP format
  - Cross-department collaboration challenges
- Current project improvements:
  - Stronger coordination between IR, Registrar, and Seamless team
  - Clearer ownership of complex data elements
  - Plan to generate PDP files directly from SIS
- Laying the foundation for long-term sustainability

Calumet College – Reviving PDP Participation

- Initially submitted 5 years of PDP data, but scripts were fragile and hard to maintain
- Use of new systems (e.g., Slate CRM) fragmented data across platforms
- Lack of integration between Slate and SIS (Empower) complicated updates
- Project stalled due to resource limits and lack of institutional rollout
- Current grant supports:
  - Strengthening Python/SQL scripts
  - Rebuilding historical data extracts
  - Improving integration between systems
  - Laying groundwork for sustainable PDP participation

Demonstration of Tools

Demonstration of Tools – Overview

- Designed to reduce institutional burden and manual effort
- Tools automate core IR functions using campus-specific inputs
- Anchored in shared data standards (e.g., CDD)
- Improve consistency, data quality, and reuse across reports
- Live demos will show real examples and integration options

Common Data Dictionary (CDD) in Action

- Foundation for shared understanding of institutional data
- Aligns definitions across institutions for consistency
- Supports standardized reporting and peer benchmarking
- Enables tool automation by referencing common data elements

ender	DOB	HS_Graduati	PreviousDual	First_Gen	Race_Ethnici	CampusLoca	CurrentYear	Term
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2021	Fall
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2021	Spring
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2022	Fall
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2022	Spring
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2024	Fall
nknown	11/19/98	Graduated	Yes	At least one	American Inc	Remote	2024	Spring
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2021	Fall
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2022	Fall
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2022	Spring
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2023	Fall
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2023	Spring
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2024	Fall
ale	6/26/83	Graduated	No	At least one	Native Hawa	Main Campu	2024	Spring
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2021	Fall
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2021	Spring
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2022	Fall
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2022	Spring
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2023	Fall
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2023	Spring
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2024	Fall
nknown	1/12/98	Partial Credit	No	At least one	Nonresident	Remote	2024	Spring
nknown	9/2/77	GED	No	At least one	White	Main Campu	2021	Fall
nknown	9/2/77	GED	No	At least one	White	Main Campu	2021	Spring
nknown	9/2/77	GED	No	At least one	White	Main Campu	2022	Fall
nknown	9/2/77	GED	No	At least one	White	Main Campu	2022	Spring
nknown	9/2/77	GED	No	At least one	White	Main Campu	2023	Fall
nknown	9/2/77	GED	No	At least one	White	Main Campu	2023	Spring
nknown	9/2/77	GED	No	At least one	White	Main Campu	2024	Fall
nknown	9/2/77	GED	No	At least one	White	Main Campu	2024	Spring

CDD Reporting Tool – Program Summary

- Built on the Common Data Dictionary (CDD) framework
- Generates standardized reports (e.g., by program, term, cohort)
- Designed to pull from various institutional data sources
- Uses aligned definitions for consistency and comparability
- Reduces manual reporting effort and duplication
- Supports transparency and campus-wide understanding

Enrollment Report		
Program	Enrollment	Surplus/Defecit
Computer Science	799	-3%
English	873	6%
Math	802	-3%
Geology	839	2%
Business Admin	818	-1%
Animation	803	-3%
Digital Media and Communications	821	0%
Fine Art	802	-3%
Certificate: Video Editing	840	2%
Certificate: Google Data Analytics	748	-9%
Certificate: Google Project Management	871	6%

IPEDS Enrollment Reporting Tool

- Automates generation of IPEDS Fall and 12 Month enrollment data
- Uses standardized variables from the Common Data Dictionary
- Replaces manual data pulls, spreadsheet wrangling, and hand-entry
- Improves consistency year-over-year
- Framework can be expanded to other IPEDS surveys

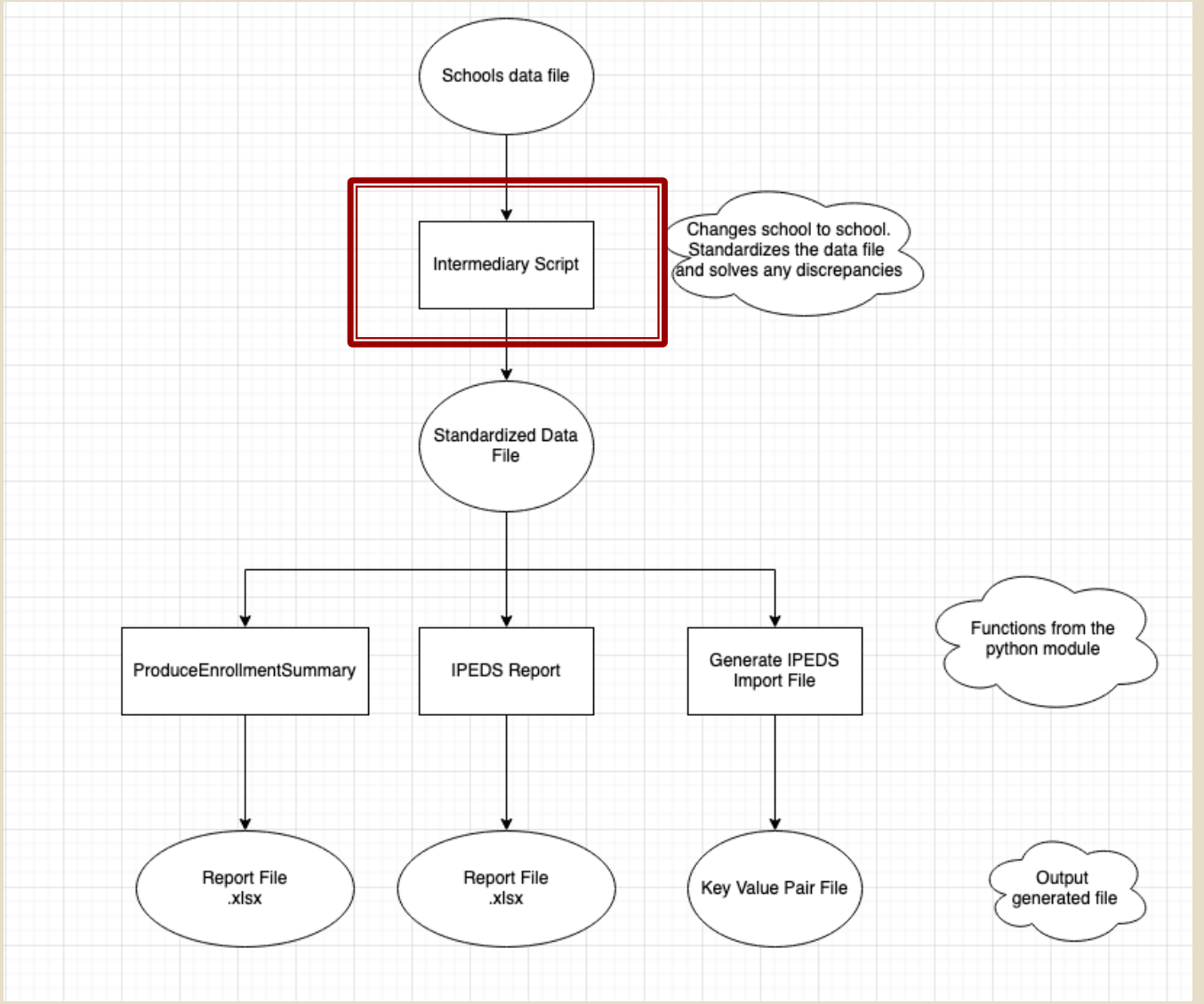
```
#!/usr/bin/python3
from reports import *
from ExcelHelpers import *

INPUT_DATA_FILE = "../../data/student_data.csv"
OUTPUT_FILE = "IPEDS_REPORT"

# Set the SD
StudentDictionary = GetStudentDictionary(INPUT_DATA_FILE)

headers = Dict_To_Array({"Author": "JoeD",
        "Date": "04/07/2025",
        "School": "Villa",
        "Fall Enrollment": "2023"
        }, False)

IPEDS2 = GetFallEnrollmentReport(dataDictionary=StudentDictionary,
                                 Year=2022,
                                 preparedbyPosition="KeyMaster",
                                 preparedbyName="Mike Smith",
                                 preparedbyEmail="Mike.Smith@villa.edu")

WriteDataToExcel(OUTPUT_FILE, headers, Dict_To_Array(IPEDS2, True))
```

PDP Data File Generator

- Converts campus data into PDP submission format
- Built using PDP specifications from NSC
- Reduces need for manual file construction or outside vendors
- Validates structure and format before submission
- Simplifies compliance with PDP data requirements

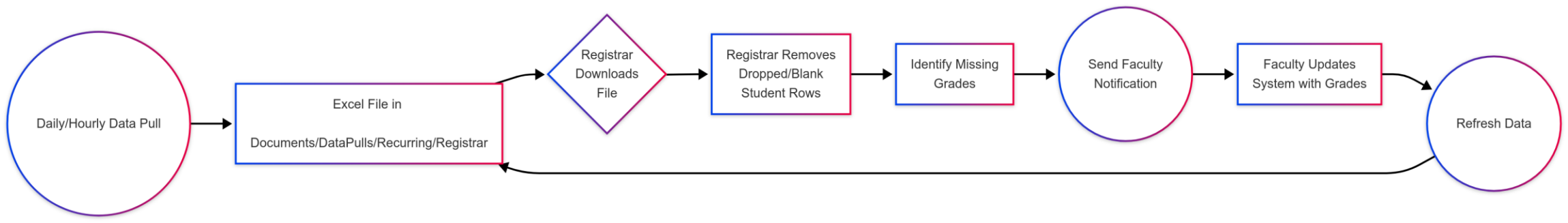
Standardizing & Automating IR Reporting with Python

- Converts campus data into desired reporting format
- Built using PDP/IPEDS/CDD specifications
- Reduces need for manual file construction
- Validates structure and format before submission
- Simplifies compliance with data submission requirements

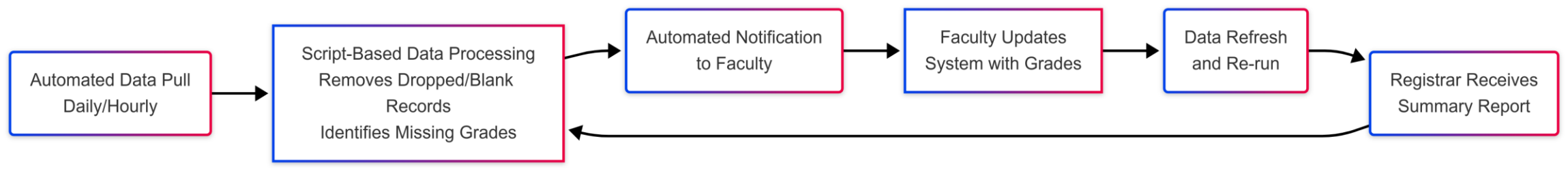


Workflow Automation

Current State – Manual Data Call



Future State – Automated Data Call



Technical Setup & Integration

- Tools are localized and run entirely on campus
- Built using basic Python libraries; requires minimal setup
- No external data transfer—writes securely to local file system
- Easily integrates with SIS and other campus databases
- Optional intermediary scripts can resolve data discrepancies
- Designed with privacy, control, and simplicity in mind

Audience Prompt

- Could you see any of these tools fitting into your institution's workflow?

Year 2 of Grant

YES WE MUST COALITION
COLLEGE SUCCESS FOR ALL

The next phase

- Data Returned from NSC
- Create Relevant Visuals
- Do Peer Comparisons
- Interpret Data
- Explore Common Data Literacy Programs
- Explore Common Technology Solutions – e.g. Data Warehouse
- Reports to Gates

Open Dialogue & Participation Opportunities

Invitation to Participate & Share

- A formal YWMC IT network will launch by November 2026
- We plan to seek additional funding to expand participation
- Would your institution be interested in joining the next phase?
  - Yes
  - No
  - Maybe
  - Need More Information
- What ideas or priorities should we highlight in the next grant proposal?