

# Data Management Plan (DMP) Incremental Credentialing



William E. Pate  
Ad Hoc Analytics  
Chevy Chase, MD

Kirk Knestis  
Evaluand, LLC  
Reston, VA

May 2023

## APPENDIX E: DATA MANAGEMENT PLAN (UPDATED MAY 2023)

---

Introduction	2
1. Types of data to be shared	2
2. Procedures for managing and for maintaining confidentiality of data	3
3. Roles and responsibilities of project or institutional staff	5
4. Expected schedule for data sharing	5
5. Format of the final electronic dataset	6
6. Documentation to be provided	6
7. Method of data sharing	6
8. Specific conditions for data sharing	7
9. Circumstances that prevent all or some of the data from being shared	8

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305T210063 to SUNY Empire State College. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

## Introduction

The Credential As You Go project will conduct a large-scale Design and Development Research study, to assess the promise of the Incremental Credentialing Framework to realize beneficial outcomes for all learners in higher education. Research will be coupled with (1) extended rapid prototyping of incremental credentials across three states (CO, NY, and NC), which includes five state systems (community college and university) and a minimum of 20 institutions; (2) inquiry as to what key elements need to be in place for institutions to adopt incremental credentialing policies and practices; and (3) identifying key elements for the design and execution of a national campaign for building awareness of and support for restructuring the U.S. postsecondary credentialing system.

The project team is committed to working in an ongoing fashion with our IES Program Officer for post-award technical assistance with details of a data sharing plan to support this project. We also understand that the Data Management Plan (DMP) is a living document that requires regular review and updates. This will be the responsibility of the PI, overseeing those tasks by key research personnel and our subcontractor.

### 1. Types of data to be shared

Qualitative evidence for the feasibility study portion of the research will include recordings (audio and video) of informant contributions to web-mediated focus groups and interviews (with participant assent), and transcriptions and notes from those sessions (as Google Doc or MSWord files for analysis purposes). Focus group input from students to inform the survey design will produce the same types of evidence. Qualitative data to be shared will be distilled to text-only documents (i.e., transcribed from audio or video files manually or with the assistance of technology), with potentially identifiable personal information redacted from the narrative record.

Additional implementation data will be collected through student questionnaires, which will include quantitative and limited created-response data reflecting respondents' understanding and value of incremental credential options available to them; and self-report immediate outcomes (e.g., transfer, some factors related to continuation, and anticipated employment results) to support institutional data. Questionnaire response data will be deidentified as necessary before sharing.

Institutional data for outcome analyses will include *Personal Information* as defined by the SUNY Empire State College [Enterprise Data Classification Policy](#), including participating individuals' names, email addresses, phone numbers, transcript information, academic records, demographic information, financial aid application and status data, IP addresses, browser and computer information, data from interactions with project websites and electronic communications, and information collected using online student self-report questionnaires.

Further, limited demographic data will be collected to analyze equity of outcomes considering key student subpopulations, including ethnicity and health/disability variables (e.g., to understand if outcomes are equitably distributed among student groups). Specific institutional outcome measures will include variables relating to student postsecondary academic success—*access, enrollment, persistence, progress, completion, transfer*, and limited information about *continuing education and employment through self-reported data*, as follows.

## Key Measures

- Access – Evidence of readily available information on, advising for, and the ability to register for targeted credentials.
- Enrollment – Registration for incremental credential, acceptance of all associated charges (i.e., one who is registered but not paid is not enrolled), and remains enrolled past the add/drop period at the institution. For comparing student data with non-participating institutions, enrollment will be defined using the IPEDS definition.
- Persistence – Term-to-term continued enrollment (per the previous definition) towards an educational goal (e.g., incremental credential, certificate or degree completion) or completion of that goal.
- Progress – Term-to-term completion of enrolled courses and/or successful attainment of credential requirements towards an educational goal (e.g., incremental credential, certificate or degree completion) or completion of that goal.
- Completion - The attainment (or rate of attainment) of formal awards (e.g., incremental credential, certificate, degree) by a student, within a stipulated period of time.
- Transfer - A transition between postsecondary institutions in which the destination institution grants the student credit for courses taken at the origin institution; normally a one-way transition (i.e., temporary enrollment at a new institution with return to the first is not a transfer).
- Continuing Education – Enrollment in the next sequential or an additional educational credential (with or without a break) at the same institution or a different institution (e.g., incremental credential to degree, associate degree to bachelor's degree).

Note on outcome measures: the *Incremental Credentialing Framework* being revised and tested challenges traditional timelines for completion of credentials. The research team anticipates that it may be necessary to revise the time-bound indicators of the above outcomes (e.g., length of the add/drop period for enrollment) as the incremental credentialing approach fundamentally changes the timeline of course content delivery.

Particularly given the number of partner institutions, departments, and separate credentials involved in the study, quantitative data may be transferred to the project researchers in a variety of forms—Google Sheet, MS Excel, or other “flat” files (e.g., .csv). Analyses will be executed using specialized applications like SPSS, SAS, and/or scripting tools such as R or Python. Data structures for institutional data will, by necessity, have to be consistent among partners to the greatest extent possible and, regardless of the final structure selected, will follow the data dictionary developed for the research. However, allowances must be made for internal institutional requirements and data specifications, so final data schemes and translation protocols will be established collaboratively with study partners, post-award.

## 2. Procedures for managing and for maintaining confidentiality of data

Procedures for managing and for maintaining the confidentiality of Personally Identifiable Information (PII) will adhere to federal-wide standards for the protection of human subjects in research; all applicable state, local, and institutional guidance and policies. Differences among rules pertaining to partners’ institutional data will be harmonized during project startup and reflected in amendments to the DMP living document.

The SUNY Empire State College Institutional Review Board (IRB) will exercise review and oversight responsibility for human subjects’ protections, supported by authorization

agreements with partner institutions as required, consistent with SUNY Empire's [Policy and Procedures for the Protection of Human Subjects Research](#). Final details of policies and practices balancing the rights of study participants and confidentiality of study data with obligations to data sharing and transparency will be codified in iterations of this DMP. Development of informed consent processes and forms will plan for, and disclose to participants, eventual archiving and sharing of research data within IRB-enforced first principles of protecting their rights and interests.

Generally speaking, qualitative evidence (i.e., narrative records from focus groups and interviews) cannot be practically collected anonymously, but will be blinded during analysis to remove evidence of individual informants' identities and to maintain confidentiality in reporting. Questionnaires to collect data for some purposes (e.g., formative feedback to inform development of processes for implementing incremental credentials) may collect truly anonymous data without PII. Questionnaires of self-reported outcomes will ask for identifying information to allow student-level analysis of data from multiple sources, such as institutional student numbers. Where this is necessary, matching (or linkage) files will be maintained to translate student PII variables to random-character or confidential identifiers for use for analysis or data files that need to be transferred among project team members. Matching files may be managed by individual institutions or by SUNY Empire, as data sharing agreements establish, but this approach will assure that only survey data without PII is used for analysis by the subaward research partners (Evaluand LLC, Ad Hoc Analytics LLC). Institutional data, notably outcomes listed above, will be managed in a similar fashion. Ultimately, individual-level data archived for external sharing will include only the project-created random or confidential identifiers and be free of variable values that could connect data to research participants, including data that might allow "deductive disclosure" of individuals (e.g., those who might be identified by cross-tabulation of variables like race, gender, and credential program enrollment).

Organizational safeguards will be established for receiving, managing inventory for, and ensuring the currency and quality of data files provided for analysis, as well as for managing derivatives of those files (e.g., cleaned files, intermediate products of analyses). All files will be subject to standardized naming, versioning, and folder structure management practices. Directory files will be utilized to document the source, date of receipt, and other details associated with any data file. Files containing student PII will be flagged for special handling as such as they are generated by partner institutions. Personnel safeguards will also be established to manage access to data files, whether or not they include student PI.

Safeguarding of data aggregated by the research consultants will leverage cloud-based technologies that provide built-in, secure file sharing functionality (e.g., G Suite Core Services, SharePoint) that are compliant with applicable data security standards. Even for de-identified data sets within the secure Drive environment, primary data files will be segregated from working files and products of analysis. Devices used for accessing and analyzing student data are username and password protected (alternatively, by PIN/biometric means), with two-factor authentication used where available. Computer screensavers will be set to automatically lock after inactivity, requiring re-entry of a password to log back in. All computers will have antivirus software installed with both updates and scans scheduled automatically.

Project team members will follow SUNY Empire State College [Information Security Incident Response Policy](#) for any breach or accidental disclosure of data to outside parties, or if accessed by unauthorized staff or consultants, immediately upon discovery of any such occurrence. The policy states: "Any employee or college affiliate that observes or suspects a

*security incident must report the incident by submitting an incident ticketed to the Information Technology Service Desk using the ticketing system at <https://www.esc.edu/service-desk/> or by calling 1-888-HELP-009 (888-435-7009) as soon as possible. The Service Desk attendant receiving the report must contact the ISO as soon as possible.”*

In addition, in the event of any concerns of a breach, accidental disclosure, or unauthorized access of the data, an immediate report will be made to the PI who will follow procedures set forth in the policy and follow-up to ensure incident responses were resolved. Any concerns about the study will be referred to the IRB Compliance Officer.

### 3. Roles and responsibilities of project or institutional staff

Nan Travers PhD (Principal Investigator) will oversee all aspects of the research, including facilitating planning for and management of data quality, security, and sharing policies and practices; managing the sharing and storing of data; and maintaining and updating the Data Management Plan. Ashley Frank (Project Coordinator) will assist the PI in these functions.

Kirk Knestis PhD (principal consultant for the research subaward) will be responsible post-award for facilitating planning for and management of data quality, security, and sharing policies and practices. He will serve in a coordinating and quality-assurance role relating to management and retention of research data by project staff and institutional partners. In that capacity, Dr. Knestis will constitute and convene regular working sessions of a Data Management Committee, including himself, William Pate (Principal, Ad Hoc Analytics LLC; contracted statistician); PI Nan Travers; Ashley Frank (Project Coordinator); the Data Coordinator at each of the states participating in the study; and representatives of partner institutions (the latter on an ad hoc basis as required to coordinate decision- and policy-making efforts relating to the study). Dr. Knestis will not have signatory authority for the grantee or any of the institutions or systems participating in the research. His role will be facilitative rather than determinative. This organizational structure should serve to appropriately engage all of the institutions contributing to the study and meet their potentially differing expectations regarding data and data management. It should also provide a mechanism for documenting and preserving the organizational memory and continuity (e.g., against possible changes in institutional staffing) necessary to plan and implement data-sharing processes over the period required by the Department.

Each state system’s designated Data Coordinator will assist in (1) gathering required data, (2) ensuring all PII are removed and project-created random identifiers (proxy IDs) are assigned, (3) organizing the data in the format indicated in the data dictionary, and (4) sharing the data into the designated storage repository (G Suite Core Services or SharePoint) through provided secure protocols.

### 4. Expected schedule for data sharing

Schedules for data sharing will be coordinated with the timeline of grant-funded activities (particularly the national campaign to distribute information about the model being studied) and with other research dissemination activities of the project. Data will be made publicly available, considering the IES Standards for Excellence in Education Research (SEER), at least prior to publication of any findings in peer-reviewed publications, and will remain available for at least 10 years beyond the end of grant-funded activities. Data will be archived at the end of each year it is generated, allowing for time to correctly prepare files and documentation. Details of the timeline for assuring accessibility of data will be revisited

at annual project progress reviews and revised as necessary, along with any changing technical requirements or other details. If findings from project data are published after the grant has been closed out, researchers and authors associated with this project will adhere to all provisions of the DMP. A final plan for executing the DMP post-closeout will also be included in the PI's closeout report.

## 5. Format of the final electronic dataset

The formats of final data files provided for sharing will be determined by the types of data. Qualitative data (i.e., narrative transcriptions of focus group and interview meetings) will be stored in plain text files (.txt). Quantitative data (student outcomes data) will be stored in comma-separated values (.csv) and/or SPSS files (.sav, .spv, or .sps). In addition, documentation to support the use of the data and technical documentation for all instrumentation and the data dictionary will be provided in plain text files (.txt) or readily sharable portable document format (.pdf) form.

## 6. Documentation to be provided

Documentation to support other researchers to secure and effectively use data from this project will include (1) a formal study protocol, describing the purpose of the data collection, Design and Development Research study design and methodology; (2) procedures used to collect data (both qualitative and quantitative); (3) a project-length timeline of the data collection; (4) clear definition of the components of the model being evaluated; and (5) documentation of implementation fidelity and quality, including measures to assess them in future studies. Products of analyses will include (1) preliminary evidence of the feasibility of implementing incremental credentials at the institution level; (2) research results addressing meaningful, broadly recognized outcomes for postsecondary learners (including subgroups not generally successful in the completion of traditional tiered credentials); and (3) documentation of baseline characteristics of the analytic sample, appropriate for future analyses.

The project team will also maintain technical documentation for all instrumentation, as well as a formal data dictionary that details the definition of variables; data types; valid value ranges (if applicable); variable field names; and other details necessary to access, understand, and re-use data from this study.

## 7. Method of data sharing

For the first three years (during the grant period), the PI will be responsible for granting permission for accessing and using the data, and for providing access. During this period, data will be stored in a secure G Suite Core Services or SharePoint site and permission to access final electronic datasets will be provided by the PI upon completing an application for data sharing (data sharing agreement). This agreement will document information about the purpose of access, how the data will be used, and terms of agreement to share results from any findings resulting from the data sharing.

Information about data sharing and the application form will be available through the *Credential As You Go* website (to be developed through this project). The website page on data sharing will provide guidance regarding purposes of the study, privacy and confidentiality standards, limitations of the data, criteria for granting access, conditions of use, and procedures to access the final electronic datasets. The PI will obtain Digital Object

Identifiers (DOI) for public acknowledgement of the datasets and for the data to be easily found by other researchers.

At the end of the three years, and as part of the closing of the research study, the PI will deposit the data files through self-publishing with Interuniversity Consortium for Political and Social Research (ICPSR), enabling open access for replication studies. Included in the uploaded data will be a description of the study, the final datasets, documentation to support the use of the data, and technical documentation for all instrumentation and the data dictionary, and links to publications referencing the data. Information as to how to access the data will be updated on the *Credential As You Go* website giving instructions on how to access datasets and supporting materials through ICPSR.

## 8. Specific conditions for data sharing

Given the relatively conventional nature of the data to be considered for this study, the project team does not anticipate requirements to accommodate any extraordinary conditions from partners for sharing of their institutional data. The project team does however expect that formal data sharing agreements will likely be required by at least some of the partners. While the project team has secured letters of agreement in principle to share data, the specifics of final, formal agreements will be negotiated during the early weeks of the project.

With respect to data sharing, the letters of agreement state:

“We agree to collaborate with the project team and partner institutions to work toward finalizing formal agreements to collect and share de-identified administrative data to be managed in compliance with the [IES Public Access to Research Policy Part II: Public Access to Data Resulting from IES-funded Grants](#).

We understand these data will contribute to the research on how best to refine and disseminate the Incremental Credentialing Framework to improve access to, persistence in, progress through, and successful completion of postsecondary education for all students, including different race and ethnicity groups, underserved populations, learners at risk of failure, and adult learners. We further understand that data will be shared to researchers outside of the project on a restricted-use basis, consistent with the [IES Public Access to Research Policy Part II](#).

We make these commitments understanding that data shared will not contain individually identifiable information, and will include: 1) feedback from administration, faculty, staff, and employers who participated in the System-Level Steering Committee and/or were engaged in the development and implementation of incremental credentials, professional development activities, and cross-state network meetings; 2) learner outcomes data on access, enrollment, persistence, progress, and completion of incremental credential; 3) learner outcomes data on transfer, continuing education, and employment post incremental credential completion, as available; and 4) learner demographic data on incremental credential areas/disciplines, prior academic and persistence performance, age, race/ethnicity, and gender.

The aforementioned types of information will be secured from institutional data systems, through distribution of questionnaires, and by qualitative means through interviews and focus groups of stakeholders. We will work with the project research team to establish the



processes necessary to implement these data-collection strategies at our institution, as required to support evaluation of the initiative.”

## 9. Circumstances that prevent all or some of the data from being shared

At the point of developing this preliminary DMP, no institutional partner has indicated any reason to restrict sharing of data to be considered for this evaluation. However, should any specific restrictions on data sharing be established by any partner during the finalizing of the data sharing agreement process, those terms will be documented in the data sharing agreement(s) and clarified in an updated DMP. The only exception to full openness anticipated at this time is the potential need to establish and apply a cell size suppression policy, to further protect the confidentiality of students by avoiding the release of information from which individual identities might be inferred through deductive disclosure.

v.2023.0519.1058