



CALIFORNIA Data System

Staffing, Technical, and Legal Governance for the Cradle-to-Career Data System

June 2020 Cradle-to-Career Workgroup
Kathy Booth and Kathy Reeves Bracco, WestEd

Over the first quarter of 2020, planning for the state data system outlined in the Cradle-to-Career Data System Act (Act) focused on developing a list of products for the first phase of the data system, including a public dashboard and query builder tool, a process for requesting data for analytical purposes, and tools to support college planning and streamline the application process. At the May 2020 Cradle-to-Career Workgroup (workgroup) meeting, the partner entities leveraged this list of prioritized products to begin to articulate the vision and mission for the data system, make foundational decisions about the scope and purpose, and examine possible decision-making structures.

Vision, Mission, Purpose, and Scope

After the full workgroup provided feedback on draft vision and mission statements at the May meeting, a small group met and developed the following proposed text, including a list of strategic objectives that define the scope for phase one of the data system. The workgroup will vote at the June meeting on whether to adopt this wording.

Vision

The Cradle-to-Career Data System connects individuals and organizations with trusted information and resources. It provides insights into critical milestones in the pipeline from early care to K–12 to higher education and employment. It empowers individuals to reach their full potential and fosters evidence-based decision-making to help California build a more equitable future.

Mission

To be California's source of actionable data and research on education, economic, and health outcomes for individuals, families, and communities; to expand access to tools and services to navigate the education to employment pipeline.

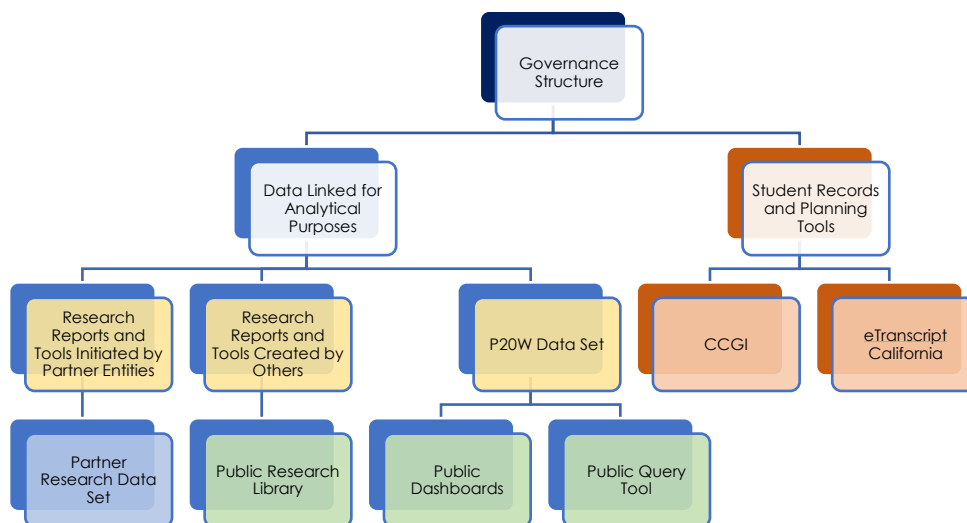
Strategic Objectives

- Develop the architecture for linking records across agencies and creating data sets
- Provide public-facing data visualizations, query tools, and a research library that provide actionable information on education, social services, employment patterns, and equity gaps in outcomes
- Provide inter-agency data sets that enable research on factors that help Californians meet critical education milestones, evaluate the long-term impact of state-funded programs, and identify strategies for closing equity gaps
- Provide resources, training, and technical assistance that build data literacy among policymakers, practitioners, and the public
- Provide college and career planning tools, college-readiness monitoring, electronic transcripts, and confirmation of eligibility for financial aid and student supports
- Lead efforts to ensure the reliability of data contributed by the partner entities

Figure 1 provides a visual representation of the strategic objectives. In addition, the Research Agenda Subcommittee is developing a document that specifies the content of the dashboard and query tool, which will be used to refine the contents of the P20W data set, based on the intended purpose of making the information available to the public. The document also identifies priority areas for research studies, which could be used to evaluate data requests.¹

¹ The current draft of the Research Agenda is posted with the materials for the June 2020 workgroup meeting at <https://cadatasystem.wested.org/meeting-information/workgroup>

Figure 1: California Cradle-to-Career Data System Phase One Components



The workgroup is now engaged in other governance questions, including the governing board structure (this topic was covered in the May workgroup background paper), staffing model, internal technical and legal framework, and external data sharing and policies (see box on this page). For each component, we highlight examples of how other states have addressed the issue, explain the current status of the issue in California, describe considerations for the workgroup, and present a set of action items and questions.

Common Components for State Data System Governance

- Mission and vision statements
- Purpose and scope statements, often in the form of a research agenda
- Decision making and staffing model
- Internal technical and legal framework
- External data sharing policies and communications

Staffing Model

As part of the governance model, other state data systems typically specify who will be responsible for maintaining and operating the state data system.

What Other States Do

In most states, the day-to-day work of the state data system is delegated to a managing entity with a director and staff. The type of managing entity and where it is housed varies by state.

In Washington, the data system is staffed by the Educational Research and Data Center (ERDC). ERDC was created by legislation and is housed within the Office of Financial Management. ERDC has a director as well as personnel who handle data warehouse development, system maintenance, and research.²

Minnesota has taken a more distributed approach. The Office of Higher Education (OHE) is the lead agency, with responsibility for managing funds, facilitating the governance process, and ensuring compliance for data security and privacy. While OHE is the lead entity and employs the state data system manager, staff from all partner entities assist with implementation on a voluntary basis.³

In both Kentucky and Maryland, the staffing models have evolved over time. When their data systems were created, both states passed laws that required participating agencies to work together and staff the state data system. Over time, as the data systems produced value to partner entities and other stakeholders—and trust increased—the states created separate agencies to house and manage the data.⁴

Current Status in California

In interviews and surveys conducted at the beginning of the planning process, there was broad consensus that a neutral third party should manage the state data system. Partner entities noted that external staffing would reduce the burden on their personnel by streamlining research requests and maintaining tools such as data dashboards and query builders. Similar time savings could be generated by leveraging staff at the California Colleges Guidance Initiative (CCGI) and eTranscript California to support students with college and career planning and to share student records as part of college and financial aid application processes. Both partner entities and advisory

² For more information on Washington's staffing structure, see <https://erdc.wa.gov/about-us/history-erdc>

³ For more information on Minnesota's staffing structure, see <http://sleds.mn.gov/#research>

⁴ Learn more about effective practices for staffing at <https://dataqualitycampaign.org/wp-content/uploads/2018/03/DQC-Cross-Agency-Gov-CaseStudy-032218.pdf>

group members emphasized the importance of the managing entity being neutral, to ensure that the state data system fulfills its stated mission and to reduce the likelihood that one partner entity would have control over or restrict access to other agencies' information.

Considerations

Given the desire for an independent party, four possible directions include:

	Example	Pros	Cons
Create a new state agency or entity	California Postsecondary Education Commission (CPEC) ⁵	<ul style="list-style-type: none"> • Greater ease in delegating legal authority to the managing entity 	<ul style="list-style-type: none"> • Governor has authority to defund • Greater cost due to overhead for an additional entity
Create a Joint Powers Authority (JPA) of the partner entities	Corporation for Education Network Initiatives in California (CENIC) ⁶	<ul style="list-style-type: none"> • Enables public agencies to jointly own the managing entity • Allows entity to accept grant funding • Flexibility in hiring 	<ul style="list-style-type: none"> • Less oversight by non-partner entities
Create a nonprofit auxiliary to a state agency or entity (such as the Office of Planning and Research)	Foundation for California Community Colleges (FCCC) ⁷	<ul style="list-style-type: none"> • Leverages existing public structures for both legal authority and overhead costs • Allows entity to accept grant funding • Flexibility in hiring 	<ul style="list-style-type: none"> • Managing entity is under the authority of a single entity

⁵ For more information about CPEC's defunding in 2011, see http://www.cpec.ca.gov/PressRelease/CPEC_Final_Release.pdf

⁶ For more information about CENIC, see <https://cenic.org/>

⁷ For more information about the FCCC, see <https://foundationccc.org/>

	Example	Pros	Cons
Create a public/private partnership (PPP) with oversight by a state agency or entity (such as between a cloud service provider and the Government Operations Agency)	Texas Online ⁸	<ul style="list-style-type: none"> Utilizes contractual safeguards signed by each agency to maintain HIPAA, FERPA, and other regulatory standards Allows for a cost recovery model 	<ul style="list-style-type: none"> Complex to arrange and maintain

However the managing entity is incorporated, the planning process has clarified the types of expertise needed, including:

- overseeing a master data management model to match individual records
- supporting a review process for data requests from third parties
- leveraging a role-based technical architecture to allow authorized parties to access specific de-identified, unitary data points
- creating data sets that are tailored to specific purposes
- providing information to the public using dashboards, query builders, and research libraries
- providing resources, training, and technical assistance
- scaling existing tools that support college planning and the transfer of student records

California has several options for how these functions could be staffed, including fully outsourcing these tasks to the managing entity or by assigning some tasks to staff from partner entities.

Suggestions to date point toward a hybrid model. For example, the Common Identifier Subcommittee recommended that the state data system underwrite staff time for each partner entity to provide expertise on data definitions and assess the quality of data matches. Members of the workgroup have suggested that managing entity staff could do a preliminary review of research requests against an approved rubric to reduce the number of proposals reviewed by partner entities, but others have clarified that staff from partner entities must support data uploads and review research methodologies

⁸ For more information on public private partnerships, see http://www.infodev.org/infodev-files/resource/InfodevDocuments_821.pdf

and conclusions. Finally, by recommending that the state data system leverage two existing tools, CCGI and eTranscript California, staffing for operational tools could be maintained within existing, publicly funded entities.

Prior Staffing Recommendations

The Education Data and Information Act of 2008 (SB 1298) authorized a planning process similar to the one being undertaken for the Cradle-to-Career Data System. In its recommendations, the governance workgroup proposed:

- The data system should be managed through a Joint Powers Authority (JPA) comprised of partner institutions, with an independent staff and budget.
 - The staff should be responsible for receiving data, maintaining a data warehouse, and ensuring privacy and confidentiality of data. Staff should also answer requests for cross-segment data, conduct research, and publish reports, as directed by the governing board.
 - The JPA should be able to contract with outside entities to expand its capacity.
-

Next Steps

Action

- Determine the structure for incorporating the managing entity
- Develop a proposed staffing framework that identifies which tasks will be led by partner entities versus the staff of the managing entity
- Based on the proposed managing entity structure, identify the membership and roles of the governing board and core committees

Questions

- Which of the four options for incorporating a managing entity would best address the intended purpose of the data system?
- Which could best provide the support for the processes identified in the technical and legal framework?

- What is the appropriate balance of responsibilities between the managing entity and the partner entities?
- For CCGI and eTranscript California, should staffing be subcontracted to the current hosts or consolidated under the managing entity?
- Should legislation be passed that codifies the staffing model?

Process

- This topic will be discussed at the June workgroup meeting, followed by two homework teams to develop proposals—one on the managing entity and staffing and one on the governing board and committee structure, for votes at the July workgroup meeting

Internal Technical and Legal Model

In order to realize its vision and mission, governance frameworks must address a wide variety of technical and legal issues such as data privacy, security, technical architecture, and information management.

What Other States Do

States create documents that codify the technical and legal frameworks for sharing information.

Technical Framework

States establish formal protocols for how data will be loaded into the system and aligned in order to produce information that can be used for dashboards, query tools, and analysis. This includes decisions about how frequently data will be uploaded and who will provide the information, such as whether information will be uploaded from a system-level entity or from individual institutions. Processes for merging data from different sources and protocols for determining matches are also part of this technical framework. Finally, in order to assure data quality and usability, the state data system

needs to establish clear documentation of all of these processes, including an inventory of all related policies, and of the definitions used for each data element.⁹

In some states, the data system is built using the architecture of one of the participating agencies, without ceding decision-making control to that agency. For example, Georgia added intersegmental information to pre-existing K–12 dashboards to make it easier for families and educators to find information on the postsecondary outcomes of their students.¹⁰ In Minnesota, the state's information technology agency is responsible for the technical infrastructure for its longitudinal early care and education data systems, while the Office of Higher Education oversees the governance for the system.¹¹

Legal Framework

Legal frameworks outlined in governance documents describe how the state data system will maintain and share data, secure student privacy, and incorporate additional entities.

States take different approaches to codifying their legal frameworks. For example, the Virginia Longitudinal Data System Book of Data Governance includes a signature page, where the heads of the participating agencies indicated their agreement with the policies and procedures, with the understanding it is subject to change. Other states, such as Washington, have all data contributors sign a Memorandum of Understanding (MOU) that establishes which data will be shared, responsibilities for data recipients, how information will be stored, and how it can be accessed.¹²

Federal and state laws shape the way that data are shared, which in turn inform the types of information in the system. For example, there are different federal rules for health and social service data governed by the Health Insurance Portability and Accountability Act (HIPAA) than for education data governed by the Family Educational Rights and Privacy Act (FERPA). State laws create further restrictions on

9 Learn more about best practices for technical and business documentation at <https://slids.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=8589>

10 Learn more about Georgia's system at <https://www.gadoe.org/Technology-Services/SLDS/Pages/SLDS.aspx> and <https://gacareerpipeline.gadoe.org/>

11 Learn more about the Minnesota system in the materials from the SLDS Learning session at <https://cadatasystem.wested.org/meeting-information/common-identifier-subcommittee>

12 Read Washington's MOU at <https://erdc.wa.gov/research-partners/our-partners/memorandum-understanding>

allowable data sharing. Most state data systems focus on education data and thus follow FERPA guidelines. Kentucky's data access and use policy provides one illustration of the kind of language in its legal framework.

KCEWS does not permit access to, or the disclosure of, any education records, workforce records, or personally identifiable information contained therein except for those purposes expressly authorized under the Family Educational Rights and Privacy Act (FERPA), Richard B. Russell National School Lunch Act (NSLA), and all other applicable state and federal data privacy laws.

Several states have articulated how additional partner entities can be integrated into existing systems. Connecticut, for example, outlines a formal process under which an agency or organization can seek approval to become a part of the state data system. The process involves a formal written request, development of a cost-sharing agreement, and unanimous approval by the chief executives of all partner entities.¹³

Current Status in California

Internal Technical and Legal Framework

Based on the recommendations made at prior workgroup meetings, the Technology & Security and Legal Subcommittees are documenting a technical and legal framework for the analytical portion of the state data system (the items highlighted in yellow, blue, and green in Figure 1).¹⁴ The framework for the operational tools (the items highlighted in orange in Figure 1) will be addressed as part of WestEd's assessment of CCGI and eTranscript California that will be provided at the July workgroup meeting.

In order to link education and social service data, the system would use a federated structure, meaning that the data would be kept in separate repositories and not stored together. Partner entities would load data into individualized, cloud-based repositories. These repositories would include the elements required for person matching and for the P20W data set at a minimum, with partner entities having discretion over any other elements that would be added.

¹³ For additional detail on Connecticut's policy, see https://www.ct.edu/files/pdfs/P20%20WIN_DataGovManual_Final_112917.pdf

¹⁴ For additional detail on the proposed framework, see the Legal and Technical Framework document included in the June meeting materials at <https://cadatasystem.wested.org/meeting-information/workgroup>

Once per year, the managing entity would link data from the individual repositories to produce:

- the P20W data set for the purpose of populating the public dashboard and query builder, which would produce summary level, de-identified data
- a research data set for the partner entities that contains unitary, de-identified data

Participation in the P20W data set would be mandated through legislation and documented through a legal agreement that could be adjusted over time to add or remove data elements. Partner entities would have discretion over which data elements to provide for the partner research data set, which would also be recorded in a legal agreement. Because records would be de-identified, each data set would be a stand-alone snapshot of the data elements approved for inclusion at the point of the annual data pull.

Data could also be linked to produce unitary, de-identified data for an approved purpose, such as a research study. While the managing entity would support the data request review process and provide the data to the requestor, each contributing partner entity must approve the request and sign a legal agreement with the requestor that defines the terms of the data access and use.

Technical Specifications and Policies

In June 2020, the California Department of Technology released a Request for Information (RFI) for a master data management solution to conduct person matching, which was developed by the Common Identifier and Technology & Security Subcommittees.

Over the summer and fall, the Technology & Security Subcommittee will be developing the following policy documents:

- Data security framework, which is a standardized series of documents and processes that define best practices to reduce cybersecurity risk
- Information governance policy, which describes how information will be collected, managed, stored, transmitted, used, reported, shared, and destroyed in a way that preserves privacy and ensures confidentiality, compliance, and security

- Data classification scheme, which clarifies the level of sensitivity for individual data elements
- Permission protocol, which identifies the level of access specific roles would have to types of data elements

Legal Documents and Policies

Over the summer and fall, the Legal Subcommittee will work on the following documents:

- Legal agreements for data sharing:
 - Recommended ways to adapt the Interagency Data Exchange Agreement (IDEA) and Business Use Case Proposal (BUCP) for partner entities
 - Agreement with the managing entity
 - Agreement with a third party to share de-identified data for an approved purpose such as a research study
- The definition for Personally Identifiable Information, which would create a consistent standard for which data elements should be subject to the highest level of privacy protections
- De-identification policy, which would create a consistent standard for how linked records would be suppressed to protect an individual's identity
- Policies on the release of records to parents, Public Records Act requests, and the exclusion of specific records related to opt-out requests
- Draft legislation to define and create the California Cradle-to-Career Data System

Considerations

In order to develop technical and legal policies and documents, the workgroup must affirm or amend the internal technical and legal framework outlined on page 10-11. The subcommittee members have requested more specificity regarding the technical and legal frameworks, so they can create specific requirements.

Technical Framework

Beyond the policies outlined above, other topics such as physical hosting requirements are best addressed once master data management options are assessed. The

Technology & Security and Common Identifier Subcommittees will be working over the summer to establish an evaluation rubric for responses to the RFI that takes these issues into account.

Legal Framework

The decision to include early care and social service participation data from the California Health and Human Services (CHHS) agency as part of the first phase of the state data system requires that the legal framework incorporate both HIPAA and FERPA, and all California laws regarding health, social service, and education data. This reinforces the value of adopting IDEA and BUCP as the legal mechanism for sharing data between partner entities, because IDEA and BUCP were initially designed for CHHS information. However, the templates need to be amended to address education-related requirements.

The Legal Subcommittee needs to resolve two additional issues:

- The Association of Independent California Colleges and Universities (AICCU) is an 501(c)(3) nonprofit association comprised of member institutions, as opposed to being a higher education system head office. Although AICCU is named as a partner entity in the Act, it does not have legal authority to provide information on individual college students to the state data system. This will need to be resolved either 1) through legislation that makes AICCU a state educational authority appointed by the state to collect information to audit and evaluate state and federal education programs for independent colleges and universities or 2) through MOUs that designate AICCU as an authorized representative, with each college delegating authority to AICCU.
- Both partner entities and advisory group members have recommended that additional agencies be included in future phases of the state data system, such as the Department of Motor Vehicles to provide information necessary to match K–12 records with employment records, or the California Department of Justice to determine whether students become incarcerated. The Legal Subcommittee could evaluate whether IDEA and BUCP can be expanded to include other types of state agencies and if a process should be established for non-state entities, such as regional data trusts.

Next Steps

Action

- Agree to an internal technical and legal framework
- Develop technical and legal policies

Questions

- Taking into account the types of information that would be available in the public dashboards and query tools, would the proposed internal technical and legal framework fulfill the stated mission of the data system?
- Does the framework conform to the California Data System rubric that was created by the workgroup?¹⁵
- How should information from independent colleges be integrated?
- How should additional partners be added?

Process

- The workgroup will vote on whether to adopt the internal technical and legal framework at the June meeting
- The workgroup will discuss the independent college and additional partner topics at the June meeting
- The technical and legal policies will be brought to the workgroup for voting at the September and October workgroup meetings

External Data Sharing Policies and Communications

The ability to stay engaged with partner entities, the legislature, and the public are key for the sustainability of state data systems, and therefore policies and processes for sharing data with third parties and communicating the value of the state data system are critical.

¹⁵ The California Data System Rubric is included on the website with the June meeting materials at <https://cadatasystem.wested.org/meeting-information/workgroup>

What Other States Do

One key aspect of state data system governance is identifying which groups can access the data and for what purposes. The text below from Michigan's Center for Educational Performance and Information (CEPI) provides one example of how states designate access.

Public: The [MI School Data](#) portal offers education data to the public in aggregate form only, and uses techniques like cell suppression for small groups to protect individual-level information.

Educators: MI School Data offers a secure login for school administrators and teachers to access their school's individual-level, non-suppressed data. Authorized users agree to comply with FERPA and all related privacy laws, as well as certifying their compliance with [CEPI Secure Report Use Policies](#). The system uses login timeouts and password expirations to help protect access.

Other State Agencies: CEPI has established formal data sharing agreements with state agencies such as the Michigan Department of Education that require access to these data. The agreements include breach procedures and liability clauses, and all parties agree to follow all laws governing the data. Records are encrypted and, where appropriate, de-identified. The data are limited to what is required for an evaluation or audit of an education program.

Lawmakers: CEPI provides federal and state required reports to lawmakers. CEPI also fills ad hoc data requests from legislators. These reports and requests use aggregate data only.

Researchers: Data are shared with researchers only for the purpose of improving outcomes for Michigan students. A designated Internal Review Board rigorously reviews all requests and monitors the process. Researchers receive de-identified data and must follow strict data storage, usage, reporting, and disposal rules. The [Michigan Education Research Institute website](#) details policies and procedures for researchers seeking data.¹⁶

States have different levels of requirements and processes for sharing data with other entities, whether the legislature or the public. Some states make general aggregated data available to the public through online dashboards, with no restrictions on access. More detailed data, which may or may not include unitary data, is typically more

¹⁶ Learn more about Michigan's data system at <https://www.michigan.gov/cepi/0,4546,7-113-985-336886--,00.html>

restricted and requires special permission to access. Kentucky, for example, has a process for requesting data that is not available through its standard reports or public use data files. Requests for aggregated data that are not publicly available do not require a data sharing agreement and typically take about three weeks to complete. Requests for individual level, de-identified data require a data sharing agreement, and the process for approval can take up to four months.¹⁷

In addition to processing data requests, Maryland uses its communications committee to ensure that the state data system adequately engages legislative staff. While most other states rely on informal mechanisms, common processes that foster communication and trust could be enshrined in a governance structure. For example, a set of guidelines could be created for collecting input from partner entity staff, practitioners, legislative staff, and the public about whether the state data system is producing useful information. Or, one of the governance committees could be tasked with ensuring there is adequate professional development to ensure a variety of audiences can confidently use publicly available data.

Current Status in California

California is designing a model that seeks to make linked data more available, as evidenced by the workgroup's endorsement of public dashboards, query builder tools, a research library, and approval-based access to de-identified unitary data. However, both the advisory groups and the community survey flagged concerns that data will not be readily accessible, particularly for research requests. The proposed federated framework, which requires partner entity approval and the collection of necessary data from partner entities, means that the state data system would be similar to the status quo, beyond establishing a consistent request process, with a designated entity tasked with shepherding the review.

Concerns have also been raised in workgroup meetings and public comment about whether there is sufficient public input about the tools that will be included in phase one, particularly given the impact of the pandemic. In response, the planning team is working with a communications firm to design a community engagement campaign for fall 2020, with support from the workgroup.

¹⁷ Learn more about Kentucky's data request process at <https://kystats.ky.gov/Reports/DataRequest>

Finally, the planning team meets periodically with legislative staff and with groups that have been advocating for a state data system, to ensure an open flow of information. Per requirements in the Act, the planning process will address training and support of users in spring 2021.

Considerations

The workgroup will need to determine how to integrate input from the general public, practitioners, the Governor's Office, and the legislature in the development of public-facing tools and the review of data requests. In addition to codifying the process for data requests from outside entities, these concerns can be addressed in the composition of the governing board and related committees and advisory groups. The workgroup will also need to consider whether it wants to craft governance guidelines for how partner entities and the managing entity will engage in public messaging.

Next Steps

Action

- Develop a request process for third parties to access de-identified, unitary data
- Create recommendations for the governing board structure, responsibilities, and membership, including communications
- Participate in fall 2020 community engagement activities

Questions

- What questions must third parties answer in the data requests process?
- How will the managing entity and the partner entities review data requests and subsequent results? Who else would be involved in this process?
- How will the state data system governing board and managing entity maintain regular communications with the partner entities, the Governor's Office, and the legislature?

Process

- A small team from the workgroup will develop a proposal for the data request process that will be voted on at the July workgroup meeting

- A small team from the workgroup will develop a proposal for the governing board structure that will be voted on at the July workgroup meeting
- A small team from the workgroup will advise on the community engagement campaign for the fall