Overview:

The P20 WIN Technical Upgrade planning group met from February to July 2021 to identify potential improvements to the technical infrastructure of the P20 WIN system. Demands on the P20 WIN system are expected to increase with the addition of new participating agencies and an increased focus on use of data to inform policy and decision-making.

The current data request cycle takes several months for a single request. A measurable goal for technical upgrades is to reduce to reduce the turnaround time for data requests, while maintaining and improving system security.

The following are the draft recommendations for review by the Data Governing Board:

1) Formalize an advisory role for DAS / BEST
2) Develop an incident response plan
3) Identify tools to streamline the data request process
4) Upgrade data transmittal applications to improve audit and notification options
5) Explore options to reduce the movement of data to increase security and improve data request management

In combination, these recommendations will improve security and increase system capacity, while preserving the core elements of the P20 WIN approach and the legal and governance process that gives agencies sovereignty and decision-making authority over their own data. The following sections describe the planning process and the rationale for the recommendations.

Review of the process:

The P20 WIN Technical Upgrade planning group met from February to July 2021 to identify potential improvements to the technical infrastructure supporting the P20 WIN system. The technical infrastructure should be responsive to increased demands and to reduce the turnaround time for requests, while ensuring security, privacy and compliance with state and federal laws and regulations.

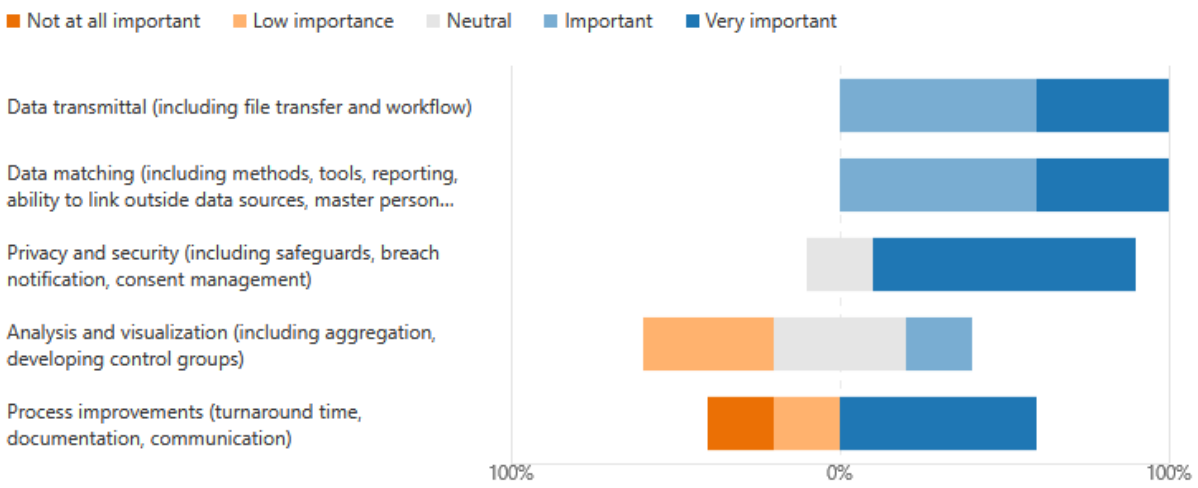
The planning process focused not only on technology, but on the people and processes that the technology supports and improvements aligned with the goals of Connecticut's SLDS grant, including:

- Adding new data sources to facilitate expansion to new agencies
- Building analytical capacity to reduce the turnaround time for data requests, particularly requests from P20 WIN participating agencies and to facilitate repeat or multi-year data requests for reporting or performance management
- Boost data matching capacity from the current environment and to manage requirements for new agencies
- Produce research and corresponding tools

The planning group membership included representatives from the P20 WIN participating agencies and from the Department of Administrative Services Bureau of Enterprise Systems and Technology (DAS / BEST) and the Office of Health Strategy (OHS) to ensure alignment with related state information technology (IT) policy and efforts. The meetings included a feedback session with 'customers' (past or potential future data requestors, from evaluators, academic researchers and agency staff) and review of

three separate use cases (current or proposed future data requests) to identify strengths and areas for improvement.

The customer feedback session identified the following topics, which the group prioritized as follows:



Recommendations:

The following five recommendations build on the existing approach for P20 WIN to support an expanded set of agencies and increased system throughput.

1) **Formalize an advisory role for DAS / BEST:** As demand on P20 WIN increases and technology evolves, P20 WIN agencies should formalize a role for a representative from DAS / BEST leadership to serve as a non-voting member on the Data Governing Board. DAS expertise and guidance has been invaluable in the planning process, and has already produced concrete improvements, including improved security settings for data transfer and identifying a potential enterprise project management and notification tool. An ongoing formal relationship will ensure benefits of this collaboration continue in the future and will allow agencies, particularly Office of Policy and Management (OPM) and Department of Labor (DOL) to benefit from the latest state resources. [Timeline / budget: Immediate / no cost]

2) **Develop an incident response plan:** P20 WIN does not have a systemwide response plan for cybersecurity incidents or data breach. DOL and OPM should work closely with the state's Chief Information Security Officer, and information security officers from participating agencies, to develop an incident response plan, accompanied by training and education for participating agency staff. Potential training topics could include: protection of human subjects, security training, etc.

   [Timeline / budget: Complete by December 31, 2021 / no cost for planning phase, implementation may have costs]

3) **Identify tools to streamline the data request process:** The current data request process is largely manual and e-mail based, for the P20 WIN Operating Group (OPM), participating agencies and data requestors. Discussion with data requestors identified 'back-and-forth' on

data requests as a major pain point for process improvements and improved technology. OPM should explore tools in the following areas:

1. **Online data request form:** The current data request form is a lengthy fillable Word document, that historically has not been available online. A brief online form would enable requestors to initiate the process more easily, which will improve the customer experience and provide a consistent format for OPM to coordinate communication between Data Recipients, the participating agencies and DOL.

2. **Improved data dictionary:** The current data dictionary is an Excel file, with a format that does not match the form for data elements required in the data request form. Improvements to the data dictionary could include:

    i. further documentation for users of the privacy, confidentiality and security requirements of individual datasets or data elements,

    ii. further documentation for users as to the meaning of data elements and

    iii. the ability to use the data dictionary to complete parts of the data request form, such as the attachments for data elements and the legal and security requirements.

    Some states have developed dynamic data catalogs or inventories that allow users to identify and select fields to develop the requisite attachments to the Data Request Form, with some of the information above.[1] The development of the data dictionary can be done in coordination with other state efforts to create consistent standards and increased transparency for metadata, such as through implementation of PA 21-35 and the State Data Plan.

3. **DSAs from template:** With system expansion, an increasing variety of data sharing agreements (DSA) will need to be generated, based on a limited set of template documents. OPM should explore tools to develop DSAs quickly and easily based on the fields and datasets selected from the data dictionary. A consistent template for P20 WIN requests may also help consistency for other interagency data requests (outside P20 WIN) and with transparency for data requestors on the legal requirements for a request.

A related change would be to explore the ability to undertake aggregate data requests to facilitate faster turnaround, although this is primarily a 'people' and 'process' change, as the existing technology already handles aggregate data requests.

[Timeline / budget: medium-term / low-cost]

4) **Upgrade data transmittal applications to improve audit and notification options:** The use cases reviewed by the planning group made clear that P20 WIN requests involve substantial movement of data, which is likely to increase with new agencies and more complex matching requirements. In addition, the enterprise agreements for P20 WIN require the Operating Group

[1] Examples include Virginia's myDATASAGE Metadata: https://metadata.cdo.virginia.gov/SAGE/DataDictionary or the Indiana MPH Data Catalog: https://hub.mph.in.gov/

(OPM) to "ensure maintenance of logs to track a) Data files received; b) matches conducted; c) Data files maintained; and d) the output files sent to approved Data Recipient(s)," including when data was transferred, who received and sent data, data Elements involved, data Sharing request to which the transferred data pertain, dates of data destruction and other relevant information. In addition, OPM and DOL will need ability to regularly monitor users with access to data to determine if continued access to data is required (for specific data requests, consistent with the data sharing agreements and confidentiality forms) and to potentially remove access.

The current data transmittal tool (Axway Secure Transport) is not configured to provide this level of information or sufficient notification for DOL or OPM staff to manage an increased volume of data requests. DOL and OPM should develop requirements, based on the planning process, to either identify adjustments to Axway Secure Transport or new tool(s) to provide the privacy, security and confidentiality information required for system audits and notification.

[Timeline / budget: Complete review by December 31, 2021 / cost TBD]

5) **Explore options to reduce the movement of data to increase security and improve data request management:** As above, the federated design of P20 WIN requires substantial movement of data, with the number of transfers increased due to the separation of matching and analytic data into separate files. While this approach is designed to reduce the risk of re-identification for individuals in the dataset, it increases the movement of data which creates security risks.

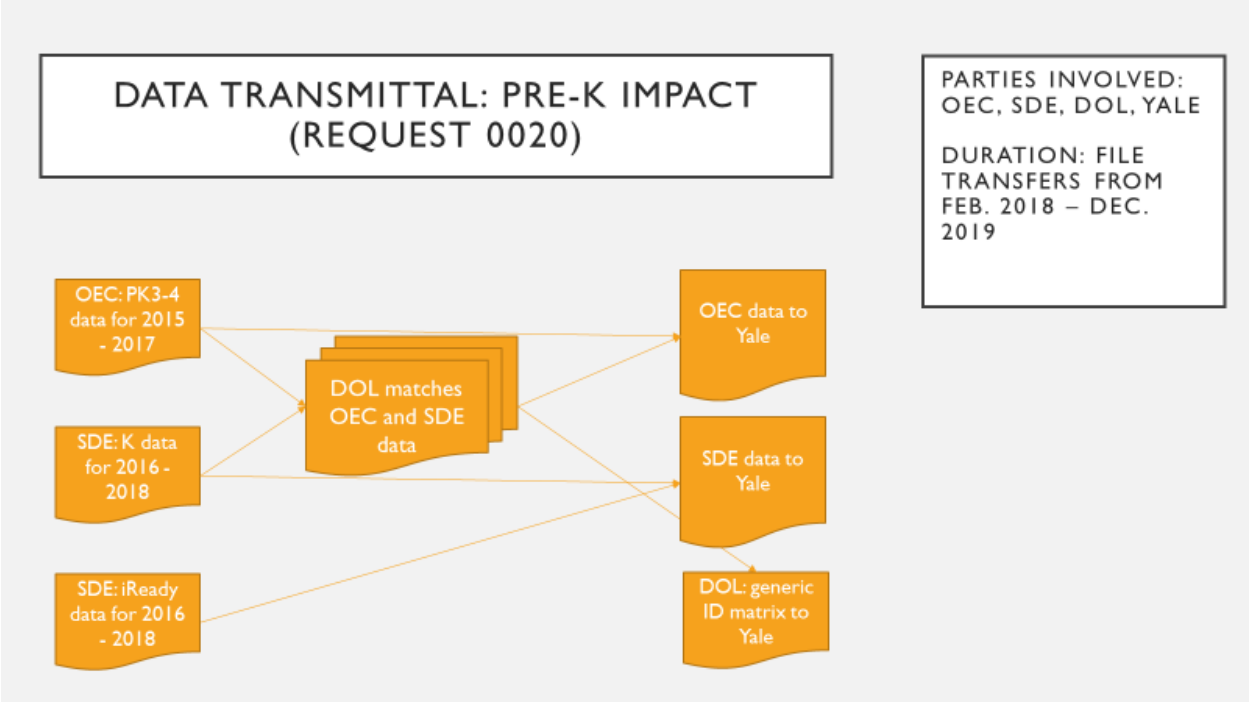
   To illustrate the current process, the schematic below maps the data transfers for one request (#0020, an evaluation of the impact of pre-K programs by the Consultation Center at Yale)[2]. The request involves the following eight distinct file transfers:

   - Match files from:
     - OEC to DOL
     - CSDE to DOL (kindergarten data)
     - CSDE to DOL (iReady data)
   - Data files from:
     - OEC to Yale
     - CSDE to Yale (kindergarten data)
     - CSDE to Yale (iReady data)
   - Generic ID matrix from:
     - DOL to CSDE
     - DOL to Yale

   Each transfer may create copies of the data, temporary files on local machines or additional copies if data are not excluded from backup procedures. The endpoints for the transfers varies, which requires coordination for data request management. A total of 10 individuals are listed as having potential access to the data for this request (one from the data requestor, Yale, and nine

[2] Documentation on the request available at: https://portal.ct.gov/OPM/P20Win/Forms-and-Agreements, the details of the data flow are listed in the 0020 Data Request – Modification 12-2019 (pdf)

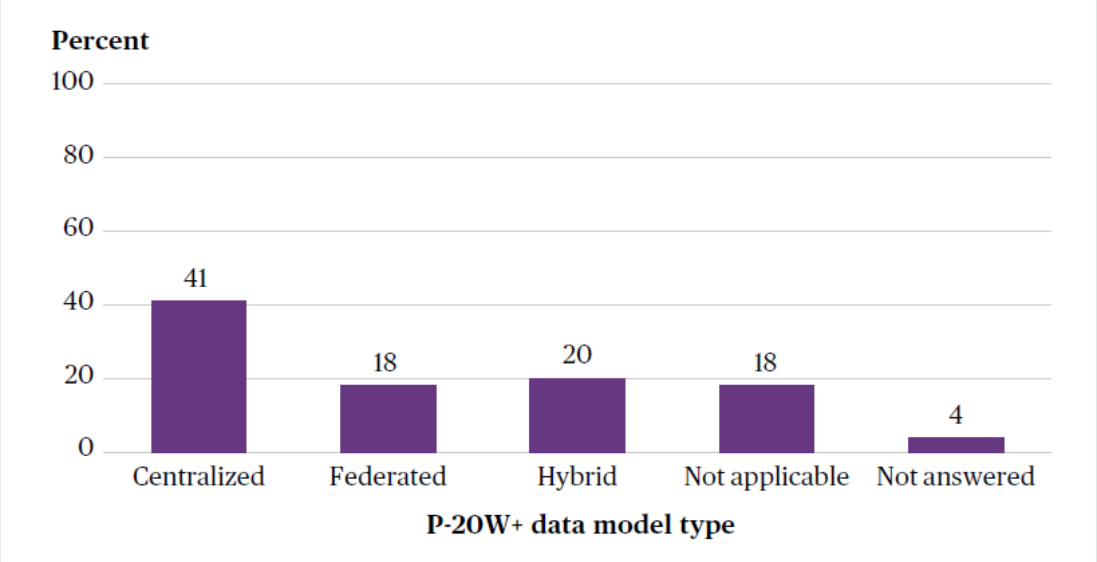
from the participating agencies). Data transfers occurred over a period of several months due to the number of datasets and revisions to the request and staffing.



The most recent national survey of SLDS shows that centralized and hybrid systems are used most often by states (Fig. 3 below) [3]. The reliance on manual transfers and processes makes P20 WIN a particularly extreme example of a federated system. A few states have pursued API-first integration (Iowa, Nebraska, Wisconsin) which can retain elements of a federated approach, in a more streamlined way.[4]

[3] Statewide Longitudinal Data Systems (SLDS) Survey Analysis: Descriptive Statistics, NCES April 2021. Federated, centralized and hybrid systems are defined on p. 7.
[4] SLDS Infrastructure Webinar: API-First System Design, IES 2019.

FIGURE 3. Percentage of states and territories with P-20W+ data collections, by model type: 2018



NOTE: N = 51. Detail may not sum to total due to rounding. P-20W+ refers to data from prekindergarten (early childhood), K-12, and postsecondary through postgraduate education, along with workforce and other outcomes data (e.g., public assistance and corrections data). The specific agencies and other organizations that participate in the P-20W+ initiative vary from state to state.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Statewide Longitudinal Data Systems (SLDS) Survey, Fall 2018.

The P20 WIN agencies should collectively explore two primary alternatives below, to determine the costs and benefits of each approach:

1. ***Review options for automating the federated model:*** Several SLDS have taken a more automated approach to the federated model. Some use application programming interfaces (APIs) or similar tools to improve security and documentation and to reduce manual steps involved in data transfer, with a mix of commercial and custom-developed solutions.[5] A similar option is to have identifiers 'hashed' by the agencies, before transfer to the Data Integration Hub. This removes the need to transmit primary identifiers, but privacy-preserving technology of this sort may require more involvement by agency staff and a consistent approach and standards for de-identification.[6] Agencies

[5] SLDS Infrastructure Webinar: API-First System Design, IES 2019.
[6] A Toolkit for Centering Racial Equity in Data Integration, AISP 2020. The Privacy-Preserving Record Linkage Pilot described on p. 58 is one example. The Administrative Data Research Facility (ADRF) uses this approach, as described in the 'Safe Data' section of their methodology: https://coleridgeinitiative.org/adrf/five-safes/#safe-data Experimental work on using privacy-preserving technology (PPT) has begun through National Center for Education Statistics (NCES) and Georgetown University: https://mccourt.georgetown.edu/news/a-federal-government-privacy-preserving-technology-demonstration/

can also develop standardized extracts that reduce the time to prepare data requests. These approaches require some costs for development or procurement to support automation.

2. ***Transition to hybrid system with integration for frequently-requested data:*** Many SLDS use a 'hybrid' approach, where frequently-requested data or key identifiers are centralized for ease of request management, and other data remain 'federated' until requested[7]. In these cases, agencies systematically identify 'use cases' for frequent or repeat reporting and then create a centralized data structure to support that limited set of use cases. A hybrid approach requires increased costs (for staffing and technology) at the central hub.

An additional approach to streamlining the hybrid model is through staffing and ***investing in a 'credentialed user' model for staffing.*** In this approach the people move, not the data. Analyst hiring for P20 WIN has started in a 'credentialed user' model, where analysts will be 'credentialed' to work on agency systems. Piloting this approach will enable agencies to determine if there are benefits for efficiency and the quality of P20 WIN results. This approach would be complementary with the above options and has the potential for cost savings through staff centralization.

A particular use case for the hybrid model is for repeat or multi-year requests that use the same data on a regular basis. Exploration of any of the options above can begin by cataloging and documenting repeat or multi-year requests to identify where the potential benefits are greatest.[8]

[Timeline / budget: medium-term / cost TBD]

Conclusion:

The list of recommendations will improve the speed, security and cost-effectiveness of the system. Improving the turnaround time for requests is necessary to have P20 WIN become a reliable and relevant resource for the state, and for the system to achieve sustainability. The recommendations will enable incremental improvements on this front, but it will take a more concerted effort and focus on the intersection of people, process and technology to see real progress.

The table below lists the many individuals that devoted time and expertise to developing the plan and recommendations.

[7] Statewide Longitudinal Data Systems (SLDS) Survey Analysis: Descriptive Statistics, NCES April 2021. Federated, centralized and hybrid systems are defined on p. 7.

[8] Actionable Intelligence for Social Policy has identified a list of frequently-used data sets for integrated data systems in their Innovation Panel Report on Establishing a Standard Data Model for Large-scale IDS Use.

Participants:

Agency	Name, Title
Connecticut Conference of Independent Colleges (CCIC)	Maura Provencher / Vice President of Research and Administration
Connecticut State Colleges and Universities (CSCU)	Jan Kiehne / Sr. Assoc. for Decision Support Resources Joe Priester / Senior Information Systems Development Manager
Connecticut State Department of Education (SDE)	Charles W. Martie / Education Consultant
Department of Administrative Service, Bureau of Enterprise Systems and Technology (DAS / BEST)	Mark Raymond / Chief Information Officer David Labbadia / Director, Application Services Debarghya Sengupta / Director of Enterprise Architecture
Department of Children and Families	Fred North / Program Supervisor, Strategic Planning, Performance Management & Evaluation
Department of Labor	Adam Hansen / IT Analyst 3 Liam McGucken / IT Analyst 2 Eric Lindquist / IT Analyst
Department of Social Services	Krithika Deepa / ITS Metrics Manager
Office of Health Strategy	Vicki Veltri / Executive Director
Office of Policy and Management	Members of Data and Policy Analytics unit

Helpful comments, edits and advice were also received from the P20 WIN Data Governing Board and Data Stewards, past and current data requestors and consultants from the SLDS State Support Team and Actionable Intelligence for Social Policy.