The Integration of Data Across Institutional Platforms to Create Comprehensive Learner Records:

Challenges and Solutions

August 2018

Data Integration Workgroup

Tom Black and V. Shelby Stanfield, Co-Chairs

Table of Contents

Introduction

During the first phase of the Comprehensive Learner Record (CLR) project (2015 – 2017), one of the most significant barriers to the development and implementation of the student record was the integration of data to populate the record with data and evidence of learning. These data, learning artifacts, assessment rubrics, catalog information and other potential material that could be included in a CLR reside in many information systems across an institution.

In the past, the sole official record of enrollment was the academic transcript. All transactions needed to create this record were collected and stored in the student information system (SIS). As institutions desire to expand the evidence of learning beyond only academic course transactions, they face challenges in collecting and joining these student data accurately and securely.

In the second phase of work that began in late 2017, a Data Integration Workgroup was assembled to identify the barriers to data integration that exist today at most institutions. While some colleges and universities will be able to address these challenges through internal resources and expertise, most lack the infrastructure to address the barriers they may face.

The AACRAO Comprehensive Learner Record (CLR) Data Integrations Workgroup is working with colleagues to provide colleges and universities with information on ways of moving forward, creating some potential guidelines and working to promote standards that not only

will benefit students but also employers and other admitting institutions (i.e., transfer, graduate, professional, licensure). The work of this group is to get others thinking broader and look to provide consistency. Part of the work will be gathering critical input from colleagues to enrich the ideas and information available to all.

The AACRAO Comprehensive Learner Record (CLR) Data Integrations Workgroup has identified the following information as components that could comprise a CLR:

⇒ Student academic record
⇒ Learning outcomes (also called skills, competencies, etc.)
⇒ Learner artifacts (dissertations, thesis, certificates, work product, etc.)
⇒ Academic program requirements, outcomes, faculty vitas
⇒ Student employment history (and associated evidence of work performance)
⇒ Student activities (role, responsibility, accomplishments, etc.)
⇒ Internships
⇒ Research activities
⇒ Service learning projects, civic engagements, honors activities
⇒ Licensures and certifications
⇒ Volunteer activities (role, responsibility, contributions)
⇒ Portfolios
⇒ Study abroad experiences and evidence of cultural competency

The manner, information systems, process, procedures, and principles by which we administer this information varies greatly between institutions; some more complete and comprehensive than others. Being at the nascent state of institutional support of the CLR, some institutions focus on these areas more than others. For those institutions who are involved (to some extent) in the CLR space, various platforms and information systems are used to collect and record multiple facets of this list, examples of which include, but are not limited to:

- ⇒ Vended Student Information System (SIS) platforms (e.g., Oracle/Peoplesoft, Ellucian Banner, Workday, Jenzabar, etc.)
- ⇒ Learning Management Systems (e.g., Canvas, Blackboard, Unizen, Desire 2 Learn, etc.)
- ⇒ Repositories for electronic dissertations, thesis, research (e.g., Scholarworks, DSpace, Fedora, Vivo, etc.)
- ⇒ Repositories for earned credentials (e.g., Badgr, Credley, NSC, Paradigm, Parchment, Credential Solutions, Digitary, etc.)
- ⇒ Curriculum Management Systems (e.g., Courseleaf, DIGARC, SmartCatalog, etc.)
- ⇒ Constituent Relationship Management Systems (CRM) (e.g., Hobsons, Salesforce, Slate, DestinyOne, Odoo, etc.)
- ⇒ Career Services (e.g., Handshake, Suitable, Simplicity, People Grove, etc.)
- ⇒ Imaging Systems (e.g., OnBase, BIS, Perceptive Software, etc.)
- ⇒ Student Organizations and Activity (e.g., AccuCampus, CampusLabs, etc.)
- ⇒ Identity Management System(s)
- ⇒ Standards Bodies (e.g., LEAP, PESC, ANSI, IMS Global, etc.)
- ⇒ Data Warehouses
- ⇒ E-Portfolios (e.g., Digication, Watermark, PebblePad, Portfolium, etc.)

Frameworks, Outcomes and Competencies

Learning occurs in a variety of places, through formal and informal environments, and can be advanced through careful consideration of the curricular structure, standards and pedagogical methods. The terminology for expressing what is to be learned is often couched in terms such as frameworks, outcomes and competencies, making more explicit the goals for education that is purposefully or intentionally provided.

One of the most recognizable credentials, the degree, is typically a formal learning structure, based on deliberate selections of courses that impart disciplinary appreciation and understanding, and in increasing frequency opportunities, the demonstration of mastery of concepts and their applicability to real-world situations. The major curriculum for a degree in this case could be referred to as a framework for acquiring particular disciplinary knowledge and developing intellectual skills.

Lumina Foundation sponsored the development of DQP (Degree Qualifications Profile), a framework that allows faculty and academic administrators to spell out what degree holders should know and be able to do. In certain degree categories, especially professional or vocationally oriented programs, e.g., Nursing and Business, clearly articulating the competencies that the degree holders have mastered can be useful to the learners and other education stakeholders. Two additional examples of well-established frameworks out of the vast number that exist include LEAP (Liberal Education and America's Promise) advanced by the Association of American Colleges & Universities that promotes the value of a liberal education through the development of intellectual and practical skills that are necessary for economic vitality and engaged citizenry, and Doing What MATTERS for Jobs and the Economy framework of California Community Colleges (CCC). Each framework articulates its values and objectives for its learners and attempts to rationalize how they are relevant temporally, economically and socially.

Commonly expressed in these frameworks are sets of learning outcomes or competencies, statements or descriptions of what is to be learned and mastered or demonstrated. These statements can be both broad and specific. For example, one of the LEAP outcomes broadly stated is Knowledge of Human Cultures and the Physical and Natural World. This learning is approached and achieved through the "study in the

sciences and mathematics, social sciences, humanities, histories, languages, and the arts"[1]. While more specifically, an example of an outcome or competency in support of the LEAP STIRS framework for developing Scientific Thinking and Integrative Reasoning Skills is to "…discuss how evidence can be used to advance knowledge and/or to inform subsequent research"[2].

While by contrast, the CCC framework emphasizes an overarching goal to "Innovate What MATTERS for Jobs and the Economy" and looks to focus education toward "investing in innovation, such as highly specialized industry training, technical consulting and a multitude of services, that solve a complex workforce training need. The end result is the ability for our system of community colleges to better deliver for employers, sectors, and their workers."[3]

It might be stating the obvious, but structure matters. Knowing what learning is being offered and for what purpose or value, and how it is being assessed and demonstrated is critical to establishing the credibility of any credential that represents that learning. According to Carol Geary Schneider, former president of AAC&U, the "…aims of education ought to be the outcomes of education."[4]

Potential Contributions from Student Employment and the Challenges of HR Systems

Experiential Learning as part of a student's undergraduate program has garnered quite a bit of attention in recent years, especially recording evidence of such experience. An area of

---

[1] https://aacu.org/leap/essential-learning-outcomes
[2] https://www.aacu.org/peerreview/2016/Fall/Riegelman , "Peer Review", Fall 2016, Vol. 18, No. 4
[3] http://doingwhatmatters.cccco.edu/Overview/DWMFramework/Innovate.aspx
[4] https://www.insidehighered.com/news/2015/08/14/aacus-moderate-strong-voice-competency-based-education-and-disruption

experiential learning warranting exploration is in the area of Human Resource (HR)/Human Capital Management (HCM) systems for institutions viewing campus student employment as a component of the CLR. These institutions seek to identify and measure student learning outcomes resulting from their employment experiences. These learning outcomes potentially relate to the domains of intellectual growth, social responsibility, leadership development, appreciating diversity, effective communication, etc.

Through institutional employment, students are provided an opportunity to be exposed to and strengthen workplace learning outcomes. Such workplace outcomes may include:

⇒ demonstrating effective stakeholder and customer-focused communication;

⇒ assessing and contributing to customer satisfaction; corresponding through oral and written word;

⇒ appreciating diversity through valuing differences from their own -- working effectively with individuals who may be different from themselves;

⇒ respecting the values and beliefs of others by seeking out others to enhance their knowledge on diversity issues;

⇒ understanding and abiding by the office's policies and procedures by being accountable for their work and maintaining a high level of accuracy and integrity with their work product;

⇒ enhancing their intellectual growth by identifying what their job entails, using critical thinking skills in problem-solving, understanding how office processes are connected with one another; and,

⇒ applying prior information to a new situation or setting; and demonstrate leadership skills by setting goals, thinking creatively to improve the quality of our services, collaborating with coworkers, and projecting a positive image of themselves and the office.

Our quick assessment is that commonly used HR systems are not as open as needed to integrate the available HR information with the other platforms comprising data included as part of the CLR.

Additional Challenges and Potential Solutions

Challenges to building the CLR fall into three main areas, two of which will be addressed in this paper. The first is developing the technology to interoperate or build the infrastructure to integrate data from these various platforms. The second main challenge is related to institutional culture and buy-in to support, prioritize, and resource the creation of the CLR. A third, learning frameworks, their governance and learning outcomes/competencies, is not addressed in this document. However, the [first phase final report](#) contains a good deal of information on this topic.

When institutions deploy new services, administrators must sort out the details. In the Information Age, these details are often a tangle of system specifications and interoperability requirements. Resident experts, when available, or outside assistants, when needed, must determine how to integrate a solution without upsetting the precarious balance of the organization's existing data ecosystem. Without proper support, this process is often

characterized as frustrating, resource-intensive, and time-consuming. It is essential to keep in mind during implementation that technology is continuously changing. Institutions must ensure there are scalable and sustainable solutions to meet ever-growing demands. Further, they must determine if a data retention plan is necessary or if archiving will be required. Also, institutions must think about whether the data is compatible or convertible as technologies emerge.

Managing competing demands and projects with limited resources is often a barrier to creating innovation at a college or university. When new projects are placed into the queue of technology projects, the demands on employee time and expertise are typically justified by campus-wide investment or an institutional or governmental mandate, and even then, Herculean efforts may be required to override resistance from employees who are asked to support existing systems while entertaining additional technology solutions. A collaborative mindset and an openness to ideation are critical elements for building a productive, successful project. A focus on the benefits to students should be maintained, especially when gaining buy-in from essential campus partners.

Information Systems Architecture, Configuration, and Design

As evident by the lists of data and respective data sources found earlier in this paper, integrating the information system data repositories that may house CLR data is critical to achieving the objectives outlined in this paper. In the early initiatives, it became apparent that there were ample data in various places within our institutions, but much of it was not in the

form that could easily be aggregated, formatted and routinely delivered without a lot of manual curation.

There are a number of approaches to architecting a CLR environment. The current prevailing design of a CLR environment makes use of an Institutional Data Store (IDS) or Electronic Data Warehouse (EDW) for which "copies" of desired data are fed from the source systems. For example, there are at least three approaches that may be considered outlined as follows;

1. Data propagated to the IDS/EDW in real-time upon updates to the source system,

2. Routine extracts from the source systems for which data are copied into the IDS/EDW based on regularly defined intervals, and

3. Data comprising the CLR are retrieved from the source systems in real-time upon access to the IDS/EDW

The manner and design of such an IDS or EDW environment is highly dependent on the interoperability and integration capabilities of the information system platforms that house the source data. For the first option, a key characteristic of the source system is being able to perform a remote procedure call (RPC) as an exit upon adding new, or updating, records to the source repository. In this case, the IDS/EDW environment will need to be "listening" for the RPC from the source system to record the new or updated records in the IDS/EDW.

The second approach typically involves the use of Extract, Transformation, and Load (ETL) routines that run at established intervals, extract the data from the respective source system,

perform any required data transformation, and send those data (via secure FTP or other method) to the IDS/EDW for loading into the CLR repository. The timing of these extracts, commonly referred to as "refresh rate", is dependent on how dynamic or frequently the source data are updated and the how critical it is for the CLR to have the most current data. The criticality of how current the information is may vary by use of the CLR and the respective authoritative data source. For example, the refresh rate from the primary student information system may need to be more frequent than from the HR/HCM system for student employee data.

The third approach noted above is highly dependent on the interoperability characteristics of both the CLR environment, as well as the source systems, and is possibly the most technically sophisticated of the three approaches. This third approach requires polling scripts to be "listening" for requests first to the CLR, then "calls" from the CLR to tasks "listening" on the source system platforms that results in the data being accessed, retrieved, and sent from the source repository back to the CLR platform for storing in the appropriate IDS/EDW table.

The underlying technologies available to perform these integrations vary by source system and their respective capabilities for sharing data. One of the most straightforward ways to load the CLR is via nightly batch extracts of data from source systems. In this typical approach, scripts run against the source systems, create a data file, and then send the file to the designated target via secure FTP. All common software platforms are capable of performing these simple data extracts and ftps.

Other means to integrate disparate platforms include web services, https calls, via specific application programming interfaces (APIs) made available by the respective system software provider, and in recent years communication software systems referred to as an enterprise service bus (ESB) have surfaced as middleware solutions to manage interactions within a distributed computing architecture. In many cases these methods of integration are available via use of open source software using established internet transport protocols and can made available from third-party software integrators.

External Partner Solutions

There are a number of organizations and consortia working in the areas of data exchange and system interoperability standards, and well as developing and making available middleware to integrate disparate platforms. Two organizations, among many, actively developing these solutions are IMS Global and DXtera.

IMS Global has a number of initiatives centered around the creation and use of standards for integrating and sharing education related data. Of particular interest is IMS Global's work developing standards referred to as Learning Tools Interoperability (LTI).

Most current generation Learning Management Systems that utilize LTI can integrate their platform into most related products. That said, LTI may be limited in the amount of data that

can be shared, so an institution may need to assess the solution of using a combination of LTI extensions along with Application Programming Interfaces (APIs) supported by the LMS. In the case of Canvas, there are APIs for most functions with available documentation (https://canvas.instructure.com/doc/api/). Such capabilities and limitations may warrant further review as CLR initiatives grow.

DXtera is a member-based consortium focusing on data integration challenges and concerns. Via their membership, a key focus of DXtera is assessing and creating "connectors" that connect legacy information systems to student applications. The goal of this approach is for providers to connect with their consumers through a series of data integration approaches. The DXtera consortium develops and shares technical assistance, education and training, and technology developments among its members.

Two organizations, Paradigm and Leepfrog, have worked together to create data integration between their platforms. This collaborative approach between companies allows the work completed in academic course management, including learning outcomes listed in Leepfrog's Courseleaf curriculum management system, to be moved into a CLR format developed by Paradigm, the Certified Electronic Diploma.

Most institutions do not have resources to provide the technologies students or other entities desire to "encompass" a student's academic history. As many institutions continue to experience budget cuts and focus on other important initiatives, the area of academic records is

left behind. Students, national organizations and/or employers may need to come together with regulations/demands or requirements to push for solution development and its critical support funding. It may be possible at this level to create an enhanced repository or repositories for student record information. This is where institutions would need to agree on some guidelines or standards. Even in these early phases of experimentation and development, it is important to consider how these data and records may be exchanged and common data elements that can be used flexibly in a future that has yet to be written.

Institutional Interest in CLRs: A Significant Integration Challenge

While early innovators participating in the first-round projects were successful in producing prototypes or early production models, it was clear that institutional commitment was a key ingredient. Participants had to be accepting, broad-based, and willing to make changes, sometimes forcing radical adjustments to organizational structures. Data collection processes and procedures, and the roles of selected personnel, later determined what technologies were needed to be successful. In these successful projects it was easy to identify at least one enthusiastic, passionate and persistent champion with enough cultural savvy to overcome resistance or to enlist support for what was required to accomplish project goals. What remains to be seen is how widespread the appetite and discipline are to expand participation in CLR.

The second challenge to building the CLR is institutional buy-in for the concept; including establishing framing principles of learning outcomes and what the records would seek to express, and the coordination needed with campus constituents who manage the respective

operations and systems. The challenge associated with the prospect of creating and offering alternative learner records is limited, or advanced, by the culture and the organization of the institution more so than any technological limitations. Cultural limitations include how decisions are made, financial priorities, the degree of buy-in by stakeholders – including the perceived academic value to expand the learner record beyond course, grade, and credit. Limitations also include the establishment of, or interface with, existing governance structures, the capacity for addressing and making innovative changes (personnel, organizational structures and know-how), and ongoing support to socializing changes and sustaining momentum.

During the first attempts at creating comprehensive learner records, some resources provided to the one dozen institutions in the project were applied to acquire or modify needed technologies in order achieve project goals. There was clear buy-in from institutional leadership for the project across most institutions in this early phase and the consistency of their leadership and their attention to the project's progress played an important role in the success and speed of the project's completion. In retrospect, this was perhaps symptomatic of early adopters rather than exposing more fundamental hurdles for being successful in creating sustainable alternative learner records.

While the concept for students having access to an electronic Comprehensive Learner Record (CLR) is attractive to many institutions, the task of creating the CLR can be daunting. As with many innovative advances that require campus coordination, shifting culture can trump any

limitations in technology as barriers to adoption. As such, what can be done to champion the creation of a CLR on these campuses?

The first step is coalescing a group to envision the consolidation of records. These may include academic history record/transcript, student outcomes and competencies as defined by faculty, extracurricular activities (i.e., student groups/Greek life), student employment history (i.e., College Work Study and off-campus jobs/internships), and co-curricular activities (i.e., research and overseas study trips). Placing these into a neat and tidy CLR takes imagination and courage. With national standards and guidelines, the CLR could someday migrate to a national cross-institutional repository. Clearly, there will be a need to exchange these data across standardized channels and link student learning to the student's educational and career paths.

The second step is to address any fears or concerns the group may have related to the creation of the CLR. Common fears are one's campus would never devote fiscal and technical resources to such an effort, the potential elimination of record-keeping staff who maintain this information and that institutional units would never share or allow access to their databases and records needed to create the CLR. With any cultural shift, there is a certain level of stepping out in faith and trust that the benefits of a new technology or innovation outweigh the fears and pains of taking the leap. When one believes the CLR to better enable our students to tell their story, and that it will be beneficial for students when seeking employment or admission to graduate school, one is armed with what it takes to impact the cultural and technical shift.

So where and how to begin? Start with an assessment of what records the institution holds and what information comprising the CLR is currently not being collected and stored. Find out how these records are maintained and create an inventory. Document the elements or records maintained, the kind of storage systems being used, and who has oversight of the records.

Next, convene a task force to explore the benefits of supporting a CLR.   Ideas for those who might comprise such a taskforce could include:

⇒ The units identified in your inventory of the records to go into a CLR

⇒ Information Technology; especially those charged with being creative and innovative and who embrace shiny, new technologies

⇒ Student leadership: both undergraduate and graduate

⇒ Marketing/Communications to help share message with stakeholders not directly involved with the project

⇒ Executive sponsor(s), such as the Provost/Senior Academic Officer or a faculty leader, to kick-off the task force and to champion the CLR as a worthwhile project.

Ask these representatives to join the exploratory task force and begin educating them on the benefits and components of a CLR. Provide them with the information available from AACRAO CLR documents. Take enough time to allow task force members to come up to speed and understand the benefits to students and to each of them as record keepers for the institution.

Share CLR success stories and examples of projects from Lumina Foundation grant-funded institutions as a way to spark creativity and ideas among your task force members.

Provide an opportunity for task force members to say why this may not be possible at your institution, and then ask the group to imagine the possibilities if money, time, and the trust that this could happen were no object. Recognize that moving to a CLR represents a huge cultural shift away from the current focus on the transcript and diploma as the only student records produced by the institution. By creating an institutionally developed and maintained CLR, students will have the opportunity to tell their story when seeking career and advanced education, validated by the institution in a way that a self-created resume, supplemented by a transcript or diploma simply cannot.

There is plenty of technology available, and much of it is available as a commodity that can be purchased, even on a subscription basis with technical support from the provider. Moreover, there are organizations interested in establishing the standards for formatting and moving data. In fact, there are at least two standards bodies (PESC and IMS Global) within American higher education that are eager to embrace new data standard requests. Finally, there is a growing array of service providers willing to modify their service offerings to accommodate emerging needs.

Many are familiar with vendors like Oracle, Microsoft, SAP, and IBM, who are currently working in the education environment to improve data integration. Additionally, there's work being

done by numerous other companies in sectors outside of education, such as healthcare and public records. Some of the better-known companies are Jitterbit, Informatica, Cerner, and Dell Boomi. It should be anticipated that some of these companies will transfer their logic and experiences over to education in attempt to address like issue.

Over the last ten years, innovative colleges and universities across the nation have deployed extensions to the academic transcript – portfolios, badges, co-curricular transcripts and others – that seek to add context to the academic record. These are rarely among the aforementioned "institutional mandates," and as such attract little interest from administrators. The traditional transcript has served students and institutions for over a century, so few see a need for adaptation. While there are many cynics who have declared higher education is insufficient, ineffective or even irrelevant, there are very few who can emphatically and empirically state what and how learning can be successfully, effectively represented. There are divided opinions within our learned academies about what educators are trying to accomplish: are they educating for the workplace or teaching how to learn, or both? This complicates the issue of what we seek to represent in a record. There are many high impact practices, i.e. internships, research, study abroad, non-curricular learning supervised and sponsored by our institutions which are not recorded anywhere, or in any explicit or expressive form, that arguably are some of the most valuable learning experienced.

To represent these learning opportunities adequately will necessarily require cultural change within our institutions. Expanding the academic record is a complex technical undertaking, but

the barriers are largely non-technical in nature. Every institution has the ability to increase the utility of its artifacts, and the registrar, through partnership with campus colleagues in academic and student affairs, has a unique ability to navigate the intricacy of our information architecture.

Data Integration – are there lessons to be gleaned from Healthcare Systems?

Student information systems are not alone in their need to bring disparate transactional and recording systems together in order to provide more robust service, insure system (and provider) interoperability and to create information repositories to support data analytics. The health care industry also faces an environment where hospitals and other health care institutions use best of breed approaches to resolve support system needs and thus Information Technology units must administer multiple electronic health record systems. This approach leads to similar interoperability issues we face in higher education. In a manner similar to our industry, these institutions must successfully transfer patient information between hospitals and agencies to enable continuity and quality of patient care. The industry also faces the same technical disruptions as our own; cloud based solutions, a plethora of startups and independent services, remote and virtual patient visits and a need to provide information and patient history across multiple platforms and for multiple users. Health care systems are as diverse as our higher education systems (if not more so), have similar privacy and security concerns and must operate at scale. Thus, it makes sense to cast an eye toward their data environment for lessons learned.

With a very broad brush, one view of the data integration in the Health Care space can be summarized in three major categories: 1) Data Standards and Protocols, 2) Open source/Ad hoc Integration tools and 3) Integration engines.

1.   Lessons from Healthcare - Data Standards and Protocols

In a manner similar to our own student information and academic record systems, there are established standards and protocols that serve to provide the fundamental means of integration. The standards can be used by each institution, but more importantly they serve as the foundation for some of the data integration tools used in the open source or middle tier environments. The common standards are governed (or at least published) by the Center for Disease Control and available at the Health Level 7 website: https://www.hl7.org/ and more information can be found on the Public Health Information Network (PHIN) site: https://www.cdc.gov/phin/. These protocols are similar to PESC and SPEEDE types of standards, but more recently these standards have evolved to include interoperability specifications called Fast Healthcare Interoperability Resources (FHIR – pronounced "fire") whose goal is to provide a simpler means to integrate data through real-time messaging APIs. This would enable the interchange of information through common and standard messaging protocols. Perhaps FHIR is a notion that should be explored to simplify higher education interoperability?

2.   Lessons from Healthcare - Open source/Ad Hoc Integration Tools

Often built on the data standards described above, there are many companies who specialize in the integration of medical information and data. Many of these companies use open source tools and platforms to accomplish the "mashing" of medical data to meet specific application and/or operational needs. The advantage of these solutions is that they are often less expensive than enterprise scale integration solutions and more easily be applied to a specific problem. However, the solutions may require either local development resources and/or contracting with consulting services to connect the new components to the integration tool and finally to the consumable product (user application, data analytic platform, etc.). Examples of smaller integration offerings in the Health Care data space are (a mention in the list is not meant as an endorsement of the company): Cypress Integration Solutions, LAIKA from MITRE, MIRTH, and Open mHealth. Note that Open mHealth is non-profit whose mission is to deploy a platform to break down the "barriers to integration and bringing clinical meaning to digital health data". The model is very similar to the Dxtera Institute, mentioned previously in this paper, in the higher education space.

3.   *Lessons from Healthcare - Integration Engines*

There are several companies that specialize in the integration of data across large and/or many systems in the health care space and in many other business spaces as well. These "integration engines" are middleware that provide data integration and system interoperability in the health care application systems. Many (if not most) are built upon the HL7 standards described above and the companies will do this at scale in two dimensions: 1) very large data repositories and 2) large number disparate operational systems. There are a number of strategies and

implementation paths for these medical institutions and corporations ranging from dynamic APIs which perform data conversions in real-time, widgets and web services that can be called dynamically, and the use of a central repository and data warehouse where data can be mashed, normalized and queried. There are many vendors in the space and many are recognizable in the higher ed business world as well: Informatica, Information Builders, Nextgen Healthcare and Snaplogic, to name a few. These solutions tend to be holistic and deployed at the enterprise level. Though we have not investigated, our assumption is that many of these solutions will be more expensive.

For a full list of the many data integration solutions, please see the following software aggregation site crated by Capterra

Since the Health care industry seems to be a step ahead of Higher Education in the world data integration, one recommendation for subsequent action would be to investigate health care data solutions more deeply and with a few case studies to determine:

1)   if Healthcare has an overarching strategy that is extensible to higher education.

2)   if there are platforms or concepts that higher education should leverage (e.g., concept of integration engine).

3)   if there are any potential partnerships and/or intersections that we can identify.

Potential Need to Categorize Data Comprising the CLR

As discussed, recently many institutions and organizations have started to look into how learning is assessed, recorded, and presented. However, as a provider of the data to others externally, the "attributes" about those pieces of data matter to those employers and hiring organizations. Many institutions have differing definitions of the components for the data. In some cases, some institutions may institutionally certify the components and others may choose otherwise. In any case, it may cause an external organization or employer to disregard information the student or the institution is able to provide. Institutions should have the ability to decide which data are certified or verified versus data that is self-reported. The larger question that should be addressed is whether the data should be more standardized or categorized globally to provide external entities the best information possible from our institutions in groupings or formats that could be readily consumed by those entities. It may be beneficial to create "buckets" or "categories" where the data can be placed. Some of these categories could be defined as achievement records, aptitudes and abilities, personal merits, comprehensive student records, co-curricular, and competency-based or e-portfolio are a few ideas. Along those lines, is looking at how institutions might better align data definitions, or at least similar characterization providing all entities more informative information. Having alignment with these categories would allow software vendors to better manage data they receive in order to provide relevant output. This would provide students being educated across multiple institutions and/or entities the ability to manage their information. Ultimately, providing employers and others consistent (or fairly standard) data and/or categories of

information from our institutions would provide more value and potentially comparable information to be utilized by others.

The information would be understood globally so that employers or hiring agencies would look at an academic transcript or the comprehensive learner record. Ideally, categories could be developed where data can be associated providing consistency and giving guidance to all of us. If institutions adopted defined categories, vendors could provide better integrations. Providing better data consistency across the higher education community would enhance the contextual understanding and relevance of the CLR. As the academic record moves toward a comprehensive learner record and it is expanded to represent comprehensive learning, having or providing guidance for developed categories where information/data can be stored would enhance what the student is conveying. This could allow many of our institutions to align learning with external agencies and employers.

Need to Digitally Link Records to the Same Learner

In addition to the technical considerations regarding system interoperability and data integration of various platforms comprising CLR data, there is also a strong need to have a well-defined machine record index to join tables connecting digital data files and records belonging to the same learner. Less than a decade ago in the US, the Social Security Number (SSN) was the primary index used to connect digital records on different files to the same student. The SSN worked well for linking or connecting records across databases for the same student given

the uniqueness of the SSN, the SSN was issued by a central agency (the Social Security Administration), the SSN was issued at the national level with each person being issued only one SSN, and the management of the SSN was tightly controlled, governed, and essentially made immutable.

A record may originate in an SIS due to any combination of documents being received by an institution over any span of time. A student record may be created in an institution's SIS due to a student attending a recruitment event, from the receipt of an application for admission, resulting from a test score received from a testing agency, a high school or college transcript being received from an institution the student previously attended, or a letter of recommendation sent to an institution on behalf of a student; just to name a few. And, it is not uncommon for these documents to be received by an institution in any sequence and over an extended period of time. In order to ensure the right documents are linked to the right student, many SIS's employ a matching algorithm to determine, that based on key data on the document received, if a new student record should be created in the SIS, or if the newly received document belongs to an existing student record in the SIS.

Historically the SSN was a critical component of the matching algorithm and worked well given the uniqueness, pervasiveness, and national presence of the SSN. As an added measure to ensure documents received "belonged" to the same student, fields such as name and date-of-birth were often used in addition to the SSN to comprise the "match-code" within the matching algorithm. In cases when the Name, SSN, and DOB matched exactly between the incoming

document and a student record already in the SIS, then the document was likely

programmatically "linked" to that student record with the SSN being the primary linking key. In

cases when the match code comprising the name, SSN, and DOB was completely unique from

any record in the SIS (i.e., did not match an existing student record in the SIS), then a new

student record would likely be created in the SIS. In cases when the name, SSN, and DOB

partially matched a record in the SIS, then the document would typically be put a suspend file

to be resolved manually. While the data and formula on which the matching algorithm was

based differed between SIS platforms and institutions, the record matching worked essentially

along these lines. This was the method in place for decades.

What has changed and thus created a problem many now face? Firstly, privacy regulations and

appropriate use of the SSN has changed significantly in recent years. In response to concerns

regarding fraud and identity theft concerns, federal and State governments have enacted laws

to limit the use and disclosure of the SSN. No longer can the SSN be required of a person unless

a federal statute exists requiring disclosure of the SSN (and if so, the federal statute must be

noted), yet if a federal statute does not exist, then providing the SSN must be optional; in which

case one must state the purpose and use of the SSN; and service cannot be withheld due to the

individual not wanting to provide their SSN. Given these restrictions on use and disclose of the

SSN, many institutions went through SSN-remediation, and thus a significant disruption

occurred in institutions' ability to effectively and programmatically match records. As a result,

institutions developed their own, institutional-specific matching algorithm and related

processes. While some of these processes are more effective than others, the result has been

increased matching errors, potential lost or mismatched documents, increased staff and manual overhead in resolving and fixing matching errors, more staff time spent searching for documents, and an increase in reputation risks with students, parents, and members of the public not understanding why their records are sometimes incomplete or with information missing from their files.

Growing interest in initiatives such as the Comprehensive Learner Record combining data from disparate platforms for the same learner, introduces the second major disruptive change. That is, with digital data belonging to the same learner yet residing in different, disparate database systems, the information used to comprise the match code used in institutions' matching algorithm may vary greatly by data type and service. Let's say the learner's name is used as a key component of the match code, we know students identify themselves by various names depending how official they may feel their interactions and engagements may be at the time they provide their name. For example, a learner may provide their full, legal name when completing an application for admission (first, middle, last, and suffix) yet may provide their nickname or what they colloquially go by when participating in a college/university recruitment event. If the name is part of the match code, then there's likely a mismatch between the learners legal first name (i.e., William) and their nickname (i.e., Billy). The same case is true with the learners' last name, as it may change due to marriage or divorce. Similarly, the time duration between engagements learners may have with an institution may introduce variations in information provided and thus used in match codes. One common element used in match codes is the learner's address; or at least permanent city, state, and zip code (either in whole or

in part). Should the student move between engagements with an institution, then different addresses may be provided on the various documents the learner provides to the institution. This too will result in records potentially not matching or requiring additional manual intervention.

What would be helpful for CLR integration is having an alternative for how the SSN had been used in the past when the SSN was the key for matching and linking records. The characteristics of such an alternative key would be it should be immutable, scalable to a national level, developed according to a set of standards and principles acceptable and adopted by all higher education services provides, systems, and platforms, and tightly governed to limit appropriate use to that of being a unique record index or key. By having a digital record key that could be used as a unique, immutable index to join tables in a distributed database network and to link/connect these records together to the single learner would be tremendously beneficial and significantly reduce lost or mismatched documents, shorten manual staff time and reduce overhead resolving matching errors, improve the student experience and confidence in the effective administration of their learner records by institutions, and allow for more efficient and effective development and deployment of CLR solutions. And, while over time institutions may have developed and refined their internal use of a unique digital record key, data portability remains a problem as information is exchanged between service providers or institutions. This again results in same potential errors in matching records and documents as well as resulting manual overhead and reputational risks. Having a scalable, national solution that provides for a unique index for matching and linking

records further enables data portability and sharing between systems, services, and institutions.

CLR Guidance in Summary

As one considers the prospect of building the CLR, the following represents the collective wisdom of individuals that either manage academic records or serve, as commercial service providers, those who manage academic records.

1. Acquire a Project Manager

Projects of the scale of CLR involve lots of stakeholders, and it is easy to get overwhelmed in this environment of exploration and uncertainty. Establish a project management approach, establish goals, and determine the criteria for determining success of the project.

2. Prototype Approaches

It isn't always clear how to represent learning that takes place in different settings and contexts. How a learning artifact will be used is an open question for many. Informed decisions are only going to be possible as you try various approaches and receive feedback regarding their value, utility and meaningfulness.

3. Perform Data Mapping Exercises

Where are the data that you wish to represent? How is it managed? What is the frequency of the updates? Are the data easily accessed? Under whose control are the data? In what form is the data? Standardized, structured, unstructured? Are the data normalized or do they come from multiple data sources?

4. Form a Data Team

Not all data are managed alike. By convening a team of data stewards, you should learn the
context of the data: how it is collected and maintained and how it is used. Work with your
team in the mapping exercises to ensure the data is consistent, accurate and properly
represented.

5. Establish Governance Around the CLR

It is important to establish the authority, either by providing the programming and/or by
assessing the learning. That authority must be legitimate and credible, defining what is being
learned, clearly stating the rigor of the learning engagement and how it is being assessed. Be
clear about what the learner has acquired or mastered as it will be translated as claims about
knowledge, skills or abilities.

6. Seek and Adhere to Applicable Data Standards

There are data standards supporting the exchange of data between institutions and
organizations; consult possible standards at: PESC.org; IMSGlobal.org; and ANSI.org.

7. Create a Digital Credential Strategy

Registrars are challenged with finding ways to support the faculty that wish to create new types
of credentials. Microcredentials, certificates, badging, lifelong learning, among others are
growing rapidly. An institutional digital credential strategy is a white paper that assesses
institutional culture, current trends and how the University Registrar envisions credentials
changing over the next few years at the institution. The three primary motivations for a digital
credential strategy are to clarify the role digital credentials play in the student lifecycle and job

marketplace; to outline the responsibilities of the Office of the Registrar in the development of these credentials; and to offer guidelines for the continued success of that development.

The Office of the Registrar must cultivate a "relational" versus a "transactional" relationship with students. Students should have access to a suite of credentials offered by the institution, and they should be considered more than customers who are party to a transaction. They purchase copies and transmissions of their credentials, and they access a platform to view and curate their credentials, but they are also involved in the creation of those credentials. They are part owner. For example, students may seek to do a non-credit internship and seek institutional backing to certify completion, job description and competencies achieved. Registrars that can provide depth to the 4-year comprehensive student experience via certified credential may create an advantage for students transitioning to the workforce.

In addition, the Office of the Registrar needs to define credential value in terms of student utility and use longitudinal data to strengthen their credential value. One way to achieve this is for Registrars to work with the Career Center and partner employers to better understand what potential employers would like to see in a credential, what learning experiences should be captured and clarity with regards to what is already being articulated. Students and their prospective employers should be able to appraise the value of the institution's credentials not only by the competencies they document but also by their completion rate (e.g. how successful are students who pursue this credential), usage rates (e.g. how often do students share or use this credential), pathway development (e.g. what types of opportunities are available to

students with this credential), and other relevant descriptors. There may be discrete metrics

that could accommodate these and other measures of value, and any used should be collected

and publicly accessible.

Appendix A. AACRAO Data Integration Workgroup Roster

V. Shelby Stanfield, University of Texas – Austin/National Student Clearinghouse, Co-Chair

Tom Black, Johns Hopkins University, Co-Chair

Mark McConahay, Indiana University – Bloomington, AACRAO Board of Directors

Frank Blalark, Duke University

Helen Garrett, University of Washington

Connie Garrick, Lone Star College

Matt Gee, University of Chicago/Brighthive

Louis Jimenez, Texas State University

Rodney Parks, Elon University

Sue VanVoorhis, University of Minnesota – Twin Cities

Randy Weber, Johnson County Community College

Tom Green, AACRAO

Appendix B. Content Contributors from Agencies and Corporations

The following individuals contributed their thoughts on and approaches to data integration at

the second Workgroup convening in Baltimore, MD, on May 30, 2018. Many of these

individuals also contributed their thoughts and ideas at the conference session about this white

paper, held at the AACRAO Tech-Transfer Conference in Minneapolis, MN, July 8-10, 2018.

Dale Allen, DXtera

Lee Brintle, Courseleaf/Leepfrog

Tom Copetto, DXtera

Doug Falk, National Student Clearinghouse

Jeff Grann, Credential Engine

Chris Jackson, Paradigm

Mark Leuba, IMS Global

Darius Robinson, Parchment