Developing a Data Integration Service for North Carolina Education Cloud

A Race To The Top Initiative

Sunday, May 20, 2012
Version 4.0
Prepared by:
Sammie Carter, Phil Emer, James Robinson, Lee Sartain
Friday Institute
{swcarter, paemer, jerobins, rlsartai}@ncsu.edu

Mark Scheible, Steve Thorpe
MCNC
{mscheible, thorpe}@mcnc.org

http://cloud.fi.ncsu.edu
cloudhelp@ncsu.edu
This page is left blank intentionally
Developing a Data Integration Service

About the Authors

Sammie Carter

Sammie Carter is a Systems Architect for the NC Education Cloud Program. He is currently leading the Data Integration and Identity and Access Management projects, and a major contributor to other RttT Initiatives. Prior to his work on the NC Education Cloud, Sammie started at the Friday Institute in summer of 2005 where he designed and built the Friday Institute's overall technology infrastructure. He served as Associate Director of Technology where he managed all IT Operations. Sammie holds a Bachelor's of Science degree in Computer Science from the College of Engineering at N.C. State University. He also serves as an Adjunct Professor in the Department of Computer Science at N.C. State.

Phil Emer

Phil Emer has spent more than 20 years working at the intersections of data communications, research, academia and business – splitting time between the public and private sectors. In the private sector Phil worked as an engineer with IBM, as a technology executive with venture-backed Carolina Broadband and as a technology director with not for profit MCNC. In the public sector Phil directed voice, video, and data communications at NC State University and now serves as Director of Technology Planning and Policy at the Friday Institute where he leads the NC Education Cloud initiative funded under the NC Race to the Top grant.

Phil serves on the NC eLearning Commission, is a member of the advisory board of the Institute for Next Generation IT Systems, and serves as adjunct faculty in the Department of Computer Science. In November 2009, Governor Bev Perdue appointed Phil to the NC Innovation Council. Phil holds a B.S. in electrical engineering from Virginia Tech and an M.S. in computer engineering from NC State University.

James Robinson

James Robinson is a Systems Architect for the NC Education Cloud. He is currently supporting the design and deployment of platforms that provide for the tagging, grouping, storage, role-based access, and analysis of learning objects and instructional content. This work supports online professional development and instructional improvement system elements of the RttT grant. Prior to his involvement with the NC Education Cloud, James worked in Cooperative Extension at NCSU, where he worked on a national grant to design and build a collaborative content management and delivery system for land-grant universities across the US. James has held several other information technology positions at NCSU as a network management expert within the campus network group and as a senior technical staff member with the Centennial Networking Lab. He holds BS and MS degrees in Computer Engineering from NCSU.
Developing a Data Integration Service

Lee Sartain

Lee Sartain is the Education Policy Specialist at the Friday Institute for Educational Innovation at North Carolina State University. His expertise is at the intersection of where technology meets education policy, with a deep understanding of institutional practices at the school district and state levels. His focus on the NC Education Cloud is Data Management and Analysis. Prior to his work on the NC Education Cloud, he provided technical leadership on the NC 1:1 Learning Technology Initiative and the NC School Connectivity Initiative.

Mark Scheible

Mark Scheible has worked in the information technology field for over 30 years and is currently the Identity and Access Management Solutions Architect at MCNC in Research Triangle Park, NC. He is a member of the Identity and Access Management project team for the North Carolina Education Cloud initiative, a Race to the Top project for K-12. He was formerly the Identity and Access Manager for North Carolina State University in Raleigh, NC and worked to establish an IAM governance infrastructure as well as an IAM Roadmap for the university. Prior to NC State Mark worked as an IT Manager with Dartmouth-Hitchcock Medical Center. It was here that Mark’s interest in Identity and Access Management was first sparked, dealing with the complexities of authentication and de-authentication in a medical setting. He also had the opportunity to explore the use of 2-factor authentication and PKI through a collaborative effort with Dartmouth College. Mark started his IT career with Digital Equipment Corporation and Compaq Computers in systems management and later as a windows system administrator. Mark is a member of the EDUCAUSE Identity & Access Management Working Group. He is also the vice-chair of the NC Federated Identity Management Task Force at MCNC/NCREN. He is also an active participant of the National K-12 Federated Identity and Access Management Task Force - a collaborative group of participants from CoSN, EDUCAUSE, InCommon and Internet2.

Steve Thorpe

Steve Thorpe is a Senior Systems Programmer Analyst on MCNC’s Advanced Services Development team, where he provides research and development on cutting-edge services to be delivered to MCNC’s K-20 customer base. Steve has extensive experience in Identity Management projects within NC’s K-20 community, including collaborations with the Department of Public Instruction and two LEAs (Rockingham and Davie County Schools) in developing the world’s first SAML-based Identity Federation to span across K-20 institutions. He has a varied background in software design, implementation and delivery, distributed computing, systems administration, computer graphics and technical leadership. Steve holds B.S. and M.S. Computer Science degrees from the University at Albany.
Contents

PREFACE .................................................................................................................................................. 9
  Acknowledgements ............................................................................................................................ 10

EXECUTIVE SUMMARY .................................................................................................................... 11

PROJECT BACKGROUND .................................................................................................................. 12

DI PROJECT OUTLINE AND SCHEDULE ...................................................................................... 15
  Phase I – DI Research and Planning ................................................................................................. 15
  Phase II – Establish DI Service Model .............................................................................................. 16
  Phase III – Build and Implementation .............................................................................................. 17
  Phase IV – Full Production and Operation ...................................................................................... 18

PART I POLICY AND GOVERNANCE ............................................................................................ 19
  Policy and Governance ..................................................................................................................... 20
  DI Policy Development .................................................................................................................... 24
    Policies, Regulations, and Rules (PRR) Workflow ........................................................................ 25

PART II REQUIREMENTS AND SPECIFICATIONS ...................................................................... 26
  Requirements and Specifications Description ................................................................................. 27
  DI High-Level Design ....................................................................................................................... 28
  Data Integration and Management Platform ..................................................................................... 30
    Real-time versus Right-time ........................................................................................................ 30
  FR1 Connectivity, Adapter Capabilities, and Data Access (Data Sources and Targets) ................. 31
    FR1.1 General Notifications and Alerts ....................................................................................... 32
    FR1.2 Connect & Monitor MFTS for various data files ................................................................ 33
  FR2 Metadata and Data Modeling Capabilities .............................................................................. 34
    FR2.2 Inspection/detection of duplicate records ....................................................................... 35
    FR2.3 Consuming only required user data fields ....................................................................... 35
    FR2.4 Detect and handle changes in data between updates/consumptions (differentials) ...... 36
    FRS2.4 Specifications .................................................................................................................. 36
    FR2.5 Matching / Joining (Entity Resolution) ........................................................................... 36
  FR3 Data Transformation Capabilities ............................................................................................. 38
  FR4 Load – Data Delivery Capabilities ............................................................................................ 39
FR5 Migration, Synchronization, and Replication ................................................................. 40
FR6 Design and Development Environment Capabilities ......................................................... 41
FR7 Operational Data Store ...................................................................................................... 42
FR7.1 Core Entity Data Model .................................................................................................. 43
FR7.2 API Requirements ......................................................................................................... 43
FR7.3 ODS and CEDARS (Data Warehouse) ......................................................................... 44
NFR1 Ed-Fi Data Standard ....................................................................................................... 46
NFR2 Common Education Data Standard (CEDS) ................................................................. 47
NFR3 Schools Interoperability Framework (SIF) ...................................................................... 48
NFR4 Scale and Scope ........................................................................................................... 49
  NFR4.1 Number of Sources and Targets ........................................................................... 49
NFR5 Flexibility, Modularity, and Minimum Dependencies .................................................... 50
NFR6 High Availability and Disaster Recovery ....................................................................... 51
  NFR6.1 High Availability Design Specifications ................................................................. 51
  NFR6.1 Platform Environments ............................................................................................ 51
NFR7 Change Management and Communication/Messaging .................................................... 52
NFR8 DI System Performance ................................................................................................. 52
NFR9 DI User Accessibility ..................................................................................................... 52
NFR10 Policy Support ............................................................................................................... 52

PART III  SERVICE MODEL .................................................................................................. 53

DI Service Model Overview .................................................................................................. 54

Strategic Principles ................................................................................................................ 54

DI Service Roles ................................................................................................................... 56
  DI Policy and Governance Role ....................................................................................... 57
  DI Service Manager Role ................................................................................................. 59
  DI Data Sources Role ....................................................................................................... 61
  DI Managed Service Role .............................................................................................. 62
  DI Target Services Role ................................................................................................. 64

Project Portfolio Management ................................................................................................ 66

Service Plan Development ..................................................................................................... 68
  Requirements of Service Plan .......................................................................................... 68
Preface

The purpose of this document is to provide a comprehensive set of information surrounding the data integration issues in the North Carolina K-12 Education system and a plan for developing a Data Integration Service to address these issues. It is important to note that the K-12 Education System includes users from Pre-Kindergarten (Pre-K) to early college high school (13) groups. The beginning of this plan clearly describes foundational support roles of the DI service for the NC Education Cloud Projects and other Race To The Top Pillars. This document is separated into five major parts. Part I describes the critical need for Policy and Governance standards to provide structure, guidance, and oversight to the DI Service. Part II dives into the high-level Requirements and Specifications for the system. Part III articulates how to create a sustainable Service Model for such a system. Part IV covers the Build and Implementation part of the project, including the deployment and migration model and the development of a Technical Implementation Plan. Finally, Part V covers Full Production and Operation of the DI Service.

The target audience for this document varies depending on which section is being read. For example, any of the beginning sections describe the background of the project and are targeted to technology directors and superintendents. The Requirements and Specifications section is aimed toward technology directors and the vendor community who would potentially be the DI Managed Service provider. The services section targets technology directors, finance, and operational support personnel.

This plan is intended to serve as a living document throughout the DI project and will be changed/updated as needed. There are numerous dependencies within the scope of this project and some requirements and specifications cannot be described until other parts are completed. As such, some sections of this document are incomplete.
Acknowledgements

**DI Working Group.** The DI Working Group, which includes representatives from DPI, has been of tremendous value to this project. It has provided vast amounts of feedback and clarifications to the DI team. It has also given us a vast insight into the DPI environment that has continuously improved the DI plan.

**Vendor Information Sessions.** We would like to extend our gratitude to the many vendors that spent hours and days with us. The vendors provided an invaluable amount of information around the Data Integration spectrum and greatly informed our planning process.

**LEA Site Visits.** Several LEA districts graciously hosted members of the DI team for visits to discuss their identity management systems. Each of these systems involves some level of data integration mechanisms. These visits were in addition to the larger statewide site interviews done by the NC Education Cloud project. These in-depth sessions have been and will continue to be immensely valuable to progress. The opportunity to meet the people and understand the systems in action greatly enhanced our research.

**IAM Project Team.** The efforts put forth by the Identity and Access Management team created a well-defined plan and service model for other state based services. The documentation developed for the IAM project has been tremendously valuable guide for the DI Service plan.
Executive Summary

The NC Education Cloud Data Integration Managed Service (DI-MS) shall provide an enterprise data utility service that collects and aggregates data residing in disparate authoritative sources and provides the cloud services and LEAs a unified view of the data via an Operational Data Store (ODS). The Data Integration Service will provide an ODS, a single access point, with standardized interfaces for consuming data. The Data Integration Service will operate on three traditional data management functions: Extract, Transfer, and Load (ETL).

The Extract function will collect data from various source systems. The Transform function will perform any necessary cleaning, reformatting, standardization, and aggregation on the data elements. Finally, the Load function will load the resulting data into the ODS, specified target systems or file formats.

Overall, the DI-MS will provide a standard and reliable service to perform data movements, consolidations, and transformations needed by various cloud services and application systems. It has the potential to reduce support costs through more effective data provisioning and by enabling faster deployments of Cloud services.
Project Background

In 2010, North Carolina was awarded a 4-year, $400M Race to the Top (RttT) award through the federal stimulus program. Receipt of this grant, designed to spur public school innovation, is a key component of North Carolina’s work to continue its momentum for school improvement. The NC RttT program includes five pillars: Statewide Standards and Assessments, Great Teachers and Leaders, Data Systems to Support Instruction, Turnaround of Lowest Achieving Schools, and State Success Factors.

An important element of the State Success Factors pillar is an NC Education Cloud with infrastructure, tools, and resources to support all NC RttT initiatives. This will be a service delivery platform for modern instructional and administration support systems. It will establish a K-12 education technology infrastructure to provide cost-effective and robust services for the LEAs, and provide digital tools and resources to support all RttT initiatives. The idea is to facilitate shared services through collaborative procurements, buying rather than building services where appropriate, using pay-for-use statewide licensing arrangements, and shared support models.

Expected benefits of the NC Educational Cloud include:

- Cost savings for the LEAs through participation in a buyer’s consortium
- Increased IT efficiency through fewer servers that are utilized at a higher capacity
- Shift in emphasis from technology support to instructional support
- Improved reliability from servers hosted and managed in data centers with regular backups, formal disaster recovery plans, and stringent service level guarantees

The two foundational components of the NC Education Cloud include:

- Service Delivery Platforms
  - Learning Management Systems
  - Learning Objects Repository
  - Collaborative Tools
  - Identity Management
- Shared Infrastructure Services
  - X as a Service
  - Data Integration

RttT Technology Snapshot

<table>
<thead>
<tr>
<th>End Users (Students, Teachers, Staff, Parents)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCVPS $6M</td>
</tr>
<tr>
<td>User Devices $25.5M</td>
</tr>
<tr>
<td>LEA Network $48.5M</td>
</tr>
<tr>
<td>LMS/CMS $10.8M</td>
</tr>
<tr>
<td>Learning Objects Repository $1.7M</td>
</tr>
<tr>
<td>Collaborative Tools $4.4M</td>
</tr>
<tr>
<td>Identity Management $1.2M</td>
</tr>
<tr>
<td>Shared Infrastructure $11.2M</td>
</tr>
</tbody>
</table>

1 “NC Race to the Top”, http://www.ncpublicschools.org/RttT/
Developing a Data Integration Service

Initially targeted shared services may include:\(^3\):

- Email migrations into the cloud (e.g. Google, Microsoft)
- Web collaboration tools
- Central hosting of administrative applications
- Statewide Directory for services delivery
- Virtual desktops inside a web browser on an old PC or on a new netbook
- Unified Learning Management System and Learning Object Repository
- Managed/shared wireless services
- Network dashboards using statewide directory
- Content filtering as a service provided by the network controlled in the LEA
- Other Network services such as security, firewall, backup, and disaster recovery

The NC Education Cloud will provide a highly reliable, highly available, server infrastructure supporting the K-12 education enterprise statewide. Specifically, it will facilitate migration from LEA-hosted server infrastructure to cloud-hosted infrastructure as a service. The primary objective of the cloud is to provide a world-class IT infrastructure as a foundational component of the NC education enterprise. Moreover, the NC Education Cloud will provide for:

- Equity of access to compute and storage resources;
- Efficient scaling according to aggregate NC K-12 usage requirements;
- Consistently high availability, reliability and performance;
- A common infrastructure platform to support emerging instructional and data systems;
- Sustainable and predictable operational cost.

Robust technology infrastructure will support data-driven decision-making, for the development of and access to online instructional resources, and to transition the focus of district technical resources from infrastructure to users and instruction. Furthermore, prudent one-time investments in technology infrastructure service platforms buy down long-term IT costs, providing sustainable funding for new instructional and leadership programs.

A centralized data integration system is a foundational component for supporting current infrastructure and the continuing shift in the IT paradigm. As more services move to the cloud, a centralized and standardized way to access and provide data to these services is required for a successful cloud model.

---

Overview of Purpose

Our Vision

The NC Education Cloud Data Integration Service shall provide an enterprise data utility service that collects and aggregates data residing in disparate authoritative source systems and provides the cloud services and LEAs a unified view of the data.

The authors’ primary motivation for creating this document is to facilitate DI-related communications within and among the NC Education Cloud Team, LEAs, vendors, similar state or regional projects, and other interested parties. We welcome feedback from all potential stakeholders, as in our view it will enhance our future success.

This document provides a snapshot of the proposed Data Integration architecture, system, and service. The purpose of the Data Integration project is to address the following challenges that K-12 stakeholders and NC Education Cloud Services face:

1. LEAs and NC Education Cloud Services need access to a large set of reliable and standardized data from disparate authoritative sources
   Most of the authoritative data sources are unique systems and require custom integration to extract data. Most of the source systems also require different processes to request and acquire the source data needed. The current data transfer utility and methods have proven unreliable, creating scalability limitations.

2. DPI and LEAs don’t have resources to manage a data integration service on their own
   The resources required to offer such a data integration service do not exist in most LEAs and continued cuts to state agency budgets are hindering DPI’s ability to maintain current systems. Each LEA application is required to perform custom data integration for each educational resource that requires student or staff information.

3. Data Integration is a foundational component to future state and national education initiatives
   Each of the planned NC Education Cloud services requires a different set of data elements and integration for those elements. This need will only grow as NC continues to adapt and align with state and national mandates. A data integration platform service will be critical as NC participates and takes advantages of national initiatives such as the Shared Learning Infrastructure and Smarter Balanced Assessment Consortium.
Developing a Data Integration Service

DI Project Outline and Schedule

It is important to understand that the DI project will establish a Service for the NC K-12 community and Cloud Services. The overall project will be completed in four major phases with each phase having multiple tasks, sub-phases, and deliverables. However, the DI system/service must constantly adapt and be viewed as an on-going program with permanent resources. The scheduled timeline for the entire DI project is October 2010 to July 2014.

Phase I – DI Research and Planning
Schedule: April 2011 – April 2012 (12 months)

Major Tasks of this Phase

Problem Definition and Scope. Understand and clearly define the challenges of stakeholders and goals of the project. Having a clear understanding of the problems will allow us to set the appropriate scope. Don’t try to boil the ocean.

Building Cross-Functional Teams. Start building cross-functional heterogeneous teams to work on tasks and deliverables for the project. These teams will adapt and evolve over time. Leveraging current policy and governance bodies at DPI and identifying key members will be extremely important to get the project started with good momentum. The teams should include representatives from all organizations that will interface with the DI system/service.

Data Integration Research. A deep and wide analysis should be done to clearly understand the DI landscape. What have other organizations done with DI, what products and services are available on the market, what are the major challenges and pitfalls of large DI projects? These are some of the important questions that need to be investigated during the initial phase of the project.

Defining Data Sources and Target Services. Data Integration is all about moving and transforming data from sources to targets in an effective and appropriate manner. It is very important to understand the details and scope for both sources and targets, because it will have a direct impact on the direction, scope, and cost of the DI system/service.

Develop DI Plan. All the points above and below will be captured in a DI plan. This plan serves as a crucial communication tool in the early phases of the project. The plan is a place to capture all information about the project, intentions, and expected results. The plan will also serve as a major anchor point throughout the project and will constantly be referenced.

Review and Approval of DI Plan. The DI plan will need to be reviewed and approved by the various stakeholders before moving to the next Phase of the project.
Phase II – Establish DI Service Model
Tentative Schedule: May 2012 – September 2012 (6 months)

Major Tasks of this Phase

**Project Portfolio Management.** Coordinate with the State of North Carolina Enterprise Project Management Office and Project Management advisor to complete administrative details of the DI Project per NC General Statutes.

**Formally establish DI Policy and Governance Role for the DI Service.** Start establishing the advisory bodies (leveraging current policy and governance bodies at DPI) that will provide governance and oversight for the DI system and service. The advisory bodies will establish the business policies and standards that will be translated into data integration rules and processes in the DI system.

**Formally establish DI Service Manager.** Formally establish the DI Service Manager that will coordinate and manage the DI Service through collaboration and negotiation with the data stewards and owners of the authoritative Source Systems and Target Services.

**Formally establish Data Sources and Target Cloud Services.** Formally establish agreements with current authoritative data providers for necessary user data. Establish initial (current) target cloud services and processes surrounding their data needs.

**Develop Service Plan.** The DI Service Manager will develop a Service Plan that described the details of how plans to deploy and manage the DI Service for the DI Sources Systems and Target Services.

**Perform RFP for Products and Services.** An RFP process will be used to select the DI Managed Service Provider that delivers the enterprise data integration needs as specified.

**Review and Approval of DI Service Plan.** The DI Service plan will need to be reviewed and approved by the various stakeholders before moving to the next Phase of the project.
Phase III – Build and Implementation
Tentative Schedule: September 2012 – April 2013 (8 months)

Major Tasks and Subphases of this Phase

Subphase 1: Data Integration

Subphase 2: Operational Data Store
Phase IV – Full Production and Operation
Tentative Schedule: May 2013 – Ongoing

Major Tasks of this Phase
PART I

POLICY AND GOVERNANCE
Policy and Governance

At the end of the Race to the Top NC Education Cloud program, the responsibility for day-to-day operations and oversight of cloud services will transition to a more permanent body. The NC Education Cloud Work plan, approved by NC State Board of Education on September 5, 2011, describes that a support organization will establish and manage process, maintain shared infrastructure, manage LEA cooperative procurement, and support LEA use of shared services in a provider-neutral manner.

The NC Education Cloud team is engaged with several advisory bodies from the K-12 community and other state education organizations. These advisory bodies provide input to the final cloud implementation plan and other project plans. This process greatly enhances our understanding of the issues and validates alignment of needs and solutions.

At the highest level, the NC Education Cloud has oversight from the NC State Board of Education, NC e-Learning Commission, and the Governor’s Education Transformation Commission. In addition, the program will have LEA oversight from the RttT Education Cloud Shared Services Advisory Committee.

Working Groups have been formed to provide specific guidance to the major elements of the program in the areas of Infrastructure and Platform. Additional ad hoc groups may be formed to provide assistance with specific projects.
There are currently five advisory bodies for the DI project:

1. DI Working Group (Technical and Application Advisory)

The tasks for the Data Integration Working Group will vary depending on the project phase but can be generally stated as:

- Serve as an advisory body exchanging ideas, concerns, and feedback between DI Architects and LEA/DPI representatives during project planning
- Provide subject matter expertise about K-12 technical systems and business processes
- Define and clarify project scope based on resources and needs
- Develop/Propose policies and procedures for DI Service
- Providing continuity between DPI and Service providers

Current DI Working Group Members are:

<table>
<thead>
<tr>
<th>NCDPI</th>
<th>Karl Pond</th>
<th>Enterprise Data Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ben Comer</td>
<td>IIS Project Manager</td>
</tr>
<tr>
<td></td>
<td>Suresh Pothireddy</td>
<td>Systems Integration</td>
</tr>
<tr>
<td>FI</td>
<td>Sammie Carter</td>
<td>Systems Architect</td>
</tr>
<tr>
<td></td>
<td>Phil Emer</td>
<td>NC Education Cloud Lead</td>
</tr>
<tr>
<td></td>
<td>James Robinson</td>
<td>Systems Architect</td>
</tr>
<tr>
<td></td>
<td>Lee Sartain</td>
<td>Policy Specialist</td>
</tr>
</tbody>
</table>

Note: Karl Pond represents the Department of Public Instruction Data Management Group (DMG).

The Data Management Group (DMG) is the official data management oversight body of the North Carolina Department of Public Instruction (NCDPI). The DMG is charged with the responsibility and authority to set policy and resolve issues concerning agency data collection, management and use. The purpose of the DMG is to enable the NCDPI to meet the responsibility of providing accurate and timely data to key stakeholders.

Comprised of representation from each business unit, the DMG supports the agency's efforts to achieve the State Board of Education's (SBE's) mission, by ensuring data quality, accountability and timeliness, all of which are essential to enabling data-driven decision-making.⁴

North Carolina Department of Public Instruction (NCDPI) Data Policy Manual
http://www.ncpublicschools.org/data/management/policies/

---

⁴ http://www.ncpublicschools.org/data/management/
2. **NC Education Cloud Shared Services Advisory Committee (Policy and Priority Advisory)**

<table>
<thead>
<tr>
<th>Region 1</th>
<th>Victor Eure</th>
<th>Perquimans County Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jeff Smith</td>
<td>Pitt County Schools</td>
</tr>
<tr>
<td>Region 2</td>
<td>John Gilmore</td>
<td>Sampson County Schools</td>
</tr>
<tr>
<td></td>
<td>Steven Myers</td>
<td>Onslow County Schools</td>
</tr>
<tr>
<td>Region 3</td>
<td>Marsha Abbott</td>
<td>Vance County Schools</td>
</tr>
<tr>
<td></td>
<td>Rhonda Moses</td>
<td>Northampton County Schools</td>
</tr>
<tr>
<td>Region 4</td>
<td>Kevin Coleman</td>
<td>Cumberland County Schools</td>
</tr>
<tr>
<td></td>
<td>Tony Tuso</td>
<td>Moore County Schools</td>
</tr>
<tr>
<td>Region 5</td>
<td>Ray Reitz</td>
<td>Chapel Hill-Carboro City Schools</td>
</tr>
<tr>
<td></td>
<td>Betty Weycker</td>
<td>Winston-Salem/Forsyth Schools</td>
</tr>
<tr>
<td>Region 6</td>
<td>Jerry Shepardson</td>
<td>Charlotte-Mecklenburg Schools</td>
</tr>
<tr>
<td>Region 7</td>
<td>Darren Hartness</td>
<td>Mount Airy City Schools</td>
</tr>
<tr>
<td></td>
<td>Scott Smith</td>
<td>Mooresville Graded Schools</td>
</tr>
<tr>
<td>Region 8</td>
<td>Monty Fuchs</td>
<td>Buncombe County Schools</td>
</tr>
<tr>
<td></td>
<td>Todd Trantham</td>
<td>Haywood County Schools</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

3. **Governor’s Education Transformation Commission (Oversight and Consultation)**

   In September of 2010, Governor Beverly Perdue established the 26-member Governor's Education Transformation Commission (GETC) to provide oversight and consultation on the use of Race to the Top funds and to coordinate the use of these funds to implement the Career and College: Ready, Set, Go! Initiative across North Carolina.⁵

4. **North Carolina eLearning Commission (Guidance)**

   The eLearning Commission is dedicated to creating and promoting a collaborative online learning environment that promotes student achievement, business success, economic stability, and lifelong learning for every citizen of North Carolina.⁶

   The Commission shall be composed of up to 30 members appointed by the Governor to serve at her pleasure for terms of two years. Commission members may be reappointed for successive terms. The persons appointed to the Commission may include representatives from educational organizations and institutions, information technology providers, nonprofits, business entities, and state and local government agencies. The Governor shall appoint a Chair and two Vice-Chairs of the Commission from the membership of the Commission.⁷

5. **State Board of Education (Governance)**

   The State Board of Education consists of the Lieutenant Governor, the Treasurer, and eleven members appointed by the Governor. The Governor's appointees are subject to confirmation by the General Assembly in joint session. Eight of the appointed members represent the eight education districts of the state. Three members are appointed from the state as at-large members. The State Superintendent of the Department of Public Instruction serves as secretary and chief administrative officer of the Board. The State Board also has seven advisors (non-voting): two high school student advisors (a junior and a senior, appointments made in the junior year for two-year service) appointed by the Governor; the NC State Teacher of the Year (serves two years; overlapping terms); the NC State Principal of the Year (serves one year); a superintendent appointed by the Governor (serves one year), and a local board of education member (serves one year; held by the NCSBA Raleigh Dingman award winner).⁸

---

⁵ http://www.dpi.state.nc.us/getc/
⁶ http://www.elearningnc.gov/
⁷ http://www.governor.state.nc.us/newsitems/ExecutiveOrderDetail.aspx?newsitemID=1438
⁸ http://www.dpi.state.nc.us/stateboard/about/
**DI Policy Development**

In a DI system, a major portion of the project is the work required to develop the policies and governance for the system/service. Data Integration systems are simply *middleware* between data sources and the targets services they serve. Policies define how the data gets accessed and transformed between the sources and target services.

Policies, Regulations, and Rules (PRRs) that support and define aspects of the DI Service will be created and revised as the project moves forward. Heterogeneous, cross-functional teams of DI architects, K-12 stakeholders, and other advisory bodies will inform the development of the PRRs. The DI Service and Implementation plan, described in the later sections of the DI plan, will also help inform Policy and Governance. These PRRs will, however, be the responsibility of the Policy and Governance teams/committees to communicate and enforce. Proposed initial categories in which PRRs need to be defined are listed in Appendix G – Proposed Policies, Regulations and Rules. Each of these areas loosely aligns with the *Requirements and Specifications* and *Service Model* described in later sections of this document. These PRRs will eventually become more formalized with structure and process in the larger NC Education Cloud context. In the beginning, however, we will be proposing PRRs in response to simple questions that need to be answered in order to make data delivery decisions.
Policies, Regulations, and Rules (PRR) Workflow

Below is a proposed workflow that describes how PRRs are developed.

A need arises for a PRR to be developed for the DI Service.

DI Team reviews DI PRR need via internal discussions and with NC Education Cloud Team.

DI Working Group discusses DI PRR need via DI Email list or scheduled DI call.

DI PRR is proposed to appropriate level of Policy and Governance for endorsement and communication.

PRR is translated and incorporated into DI Service as a rule, policy, workflow, etc.
PART II

REQUIREMENTS AND SPECIFICATIONS
Requirements and Specifications Description

Any model of a systems engineering process includes activities aimed at capturing requirements. This provides a way to express what the customers and users expect the system to do. Our understanding of the intent and functions starts with an examination of requirements.

A requirement is a feature of the system or a description of something the system is capable of doing in order to fulfill the system’s purpose\(^9\).

A functional requirement [FR] describes an interaction between the system and its environment as to what the system will do. “The system shall ...”

A nonfunctional requirement [NFR] describes a restriction on the system that limits our choices for constructing a solution to the problem. “The system shall be ...”

Quality

Quality requirements define specific performance expectations that customers and users have for identified functional requirements.

Constraints

Constraints are the rules, relations, and conventions that determine the boundaries and limitations that a requirement must be realized within.

A specification restates the requirements in technical terms appropriate for the development of a system design.

There must be a direct correspondence between each requirement and specification.

Formatting and Layout:

- FR1 Functional Requirement 1
  - FRS1 Functional Requirement 1 Specification
    - FR1.1 Functional Requirement 1.1
      - FRS1.1 Functional Requirement Specification
      - Examples
    - FR1.2 Functional Requirement 1.2
      - FRS1.2 Functional Requirement Specification
      - Examples
- NFR1 Non-Functional Requirement 1
  - NRFS1 Non-Functional Requirement 1 Specification
    - NFR1.1 Non-Functional Requirement 1.1
      - NRFS1.1 Non-Functional Requirement 1.1 Specification
      - Examples
    - NFR1.2 Non-Functional Requirement 1.2
      - NRFS1.2 Non-Functional Requirement 1.2 Specification
      - Examples

DI High-Level Design

We currently have several high-level illustrations that potentially describe a proposed model for addressing the data integration challenges facing K-12 stakeholders and cloud services. These illustrations show the major components of an NC Education Cloud Data Integration System and how they could potentially interact with the authoritative data sources and various cloud-based service providers.

The following figures illustrate potential components of the desired system.
Developing a Data Integration Service

Data Integration - Basic ETL + ODS
Data Integration and Management Platform

There is a common need across all the NC Education Cloud Projects and other RttT projects for a comprehensive, enterprise data integration and management platform. A platform for extracting, migrating, syncing, matching & merging, monitoring, validating, transforming, loading, and replicating data from disparate authoritative data sources to target systems in a secure manner. The core of the integration platform is an Extract, Transform, and Load (ETL) engine but has a much larger scope for enabling consistent, accurate, and timely data throughout our organization.

The system must communicate with the authoritative data sources and retrieve the user data. User data will include numerous attributes about students, staff, and parents in the NC K-12 Education System. The user data needs to be profiled, processed, transformed, and loaded into specified target systems in a specific format and at a certain speed/frequency.

Real-time versus Right-time

Real-time data integration technology is about the speed and frequency of fetching and delivering data. Throughout this plan, we hint at the idea of real-time but what is more important is the delivery of the data in the “right-time.” Business requirements for data delivery will determine the speed and frequency -- right-time data with real-time processing.

---

FR1 Connectivity, Adapter Capabilities, and Data Access (Data Sources and Targets)

The DI System needs to provide connectivity to virtually all data sources and types, operating systems, and hardware environments using both native access and open standards. All data transferred between the source/target systems and DI must be through a secure, encrypted channel. The ability to interact with a range of different types of data structure, including:

- Native support for SQL-based processing.
- Access to static and streaming data for sending and receiving via Web Services.
- Support for unstructured and semi-structured emergent data types of less structured nature, such as email, websites, office productivity tools and content repositories.
- Support for message-oriented middleware, including WebSphere MQ from IBM, Oracle ESB, and Java Message Service (JMS).
- Schools Interoperability Framework
- Native access methods that reduce the need for custom coding.
- Specialized table loaders that optimize bulk loading of Oracle and DB2.
- Legacy and non-relational databases
- Various file formats
- Directory Services such as Active Directory, LDAP v3, Novell eDirectory
- XML
- Software-as-a-service (SaaS) and cloud-based applications and sources.
- Externalized parallel distributed processing (such as Hadoop Distributed File System [HDFS], MongoDB, and other nonSQL type repositories.

In addition, data integration tools must support different modes of interaction with this range of data structure types, including:

- Bulk acquisition and delivery
- Granular trickle-feed acquisition and delivery
- Changed-data capture (CDC) – the ability to identify and extract modified data
- Event-based acquisition (time-based or data-value based).
FR1.1 General Notifications and Alerts

The DI system must have the ability to create alerts/notifications to appropriate personnel based on system or processing events. This function must be in place across the whole DI system. Processing of the source data is the first step in the process and it is important to alert support personnel of any data file anomalies or processing errors during this step.

Example: The input user data file changes from .csv to .sql. While DI may be able to consume a .sql file, an alert/notification should be sent to administrator(s) describing that the normal .csv file was not found.

Example: If the number of incoming users records changed from 1,000,000 to 100, that would be considered a significant change and an administrator should be alerted via some method.

FRS1.1 Specifications

The DI system must be able to create alerts/notifications via email, text, pager, and web services based on administrator-defined parameters (and system parameters). The data integration tool should have an elaborate set of logging capabilities that would be able to record all activities related to monitoring, consuming and processing of the input user data. Every activity should be logged but only certain events will generate/require a notification. The notifications should be able to be customized and preconfigured based on administrator needs.

Example: If an error occurs when the data integration tool retrieves the latest user data files and attempts to unzip the archive, this action and failure should be logged and a notification should be sent to the administrators notifying them of the issue.

Example: 1,500 of 1,000,000 input records from the NCWise system had incorrectly formatted NCWiseID fields, therefore those records are considered invalid and should not be used to update the system. The administrator(s) should be notified of this issue in an easily understandable format: e.g. in a single notification or report that summarizes the 1,500 errors, rather than a flood of 1,500 individual notifications.
FR1.2 Connect & Monitor MFTS for various data files

It is expected that the DI system will not have direct access at the start of the project to some of the authoritative data source systems and must access a copy of the data from those systems through the NCDPI Managed File Transfer System (MFTS). The MFTS is essentially a secure FTP server system. On a periodic basis, the data is extracted from the authoritative sources into a tab-delimited, flat text file format, then zipped up and put into a folder on the MFTS that the DI system has access to. This task is completed by the authoritative data source systems and MFTS services. An MFTS file naming convention has been provided to the DI team to assist with the MFTS process. The NCWISE data includes approximately 1462 data fields for each record, as documented in the NCWISE data dictionary\(^\text{11}\). The MFTS file naming convention is located in Appendix H. The current data extracts field definitions for NCWISE and UID are located in Appendix I and J respectively for reference examples.

FR2 Metadata and Data Modeling Capabilities

As Metadata Management and Data Modeling become increasingly important in helping to understand data content and relationships, the DI system capabilities must include:

- Automated discovery and acquisition of metadata from data sources, applications, and other tools.
- Data model creation and maintenance.
- Physical to logical model mapping and rationalization.
- Defining model-to-model relationships via graphical attribute-level mapping.
- Lineage and impact analysis reporting, via graphical and tabular format.
- An open metadata repository, with the ability to share metadata bidirectionally with other tools.
- Automated synchronization of metadata across multiple instances of tools.
- Ability to extend the metadata repository with customer-defined metadata attributes and relationships.
- Documentation of project/program delivery definitions and design principles in support of requirements definition activities.
- Business analysts/end-user interface to view and work with metadata.

To prepare data properly, the system must perform actions like join data from multiple sources and data quality function: validate, standardize, de-dupe, enhance. The goal is to deliver clean, compliant, complete, contextual, auditable data.

Note: Only data that is needed by the target services will be profiled and modeled. The concept of least privilege will be supported. There is already a current metadata repository at DPI that can be leveraged.

---

12 Gartner Magic Quadrant for Data Integration, October 2011, ID: G00219411
FR2.1 Validate User Data

The system must have the ability to validate and cleanse user data based on pre-determined specifications/metrics. The system must be able to handle null or invalid input data records (e.g. identify incomplete or bad data fields/attributes) and alert an administrator before consuming that user data record. It must have self-protection features/capabilities to avoid consuming bad data from the authoritative data sources. Input data with a minimum set of attributes must be stored in a way that allows for various maintenance on it using the Create, Read, Update, and Delete (CRUD) model.

- **Example:** Does every user student record have a unique student identification number based on what a student id number looks like: 9-digit number, no alpha? If the data integration component checks and a student identification number is 123abc, it should create an alert/notifications for the administrator.

- **Example:** Perhaps having the ability to create pre-defined metrics that set the scope of input data and if input data does not comply/conform with those metrics then don’t consume and send an alert.

- **Example:** During processing, a student record is encountered that does not have the minimum set of fields populated with valid data (e.g. missing first name, last name, DOB, etc.), then the system should reject that record and notify the administrator.

FR2.2 Inspection/detection of duplicate records

The DI system must have the ability to detect duplicate records upon consumption of the data.

**Example:** During one of the authoritative data update cycles, the system detects two records with the same unique student identification id. If the records are identical for all the fields that are of interest to the DIS, only one is kept and the other is ignored. If the records have fields of interest that differ, the system should note the discrepancy (perhaps alerting the administrator) and be capable of either keeping one of the records, or tossing them both out.

FR2.3 Consuming only required user data fields

The DI system must have the ability to only consume pre-determined fields that define the user data set. The input data may contain non-relevant data fields that should be ignored by the system.

**Example:** The list of NCWise and UID attributes that are needed might include:

a) **Student Accounts from NC WISE:** StudentID, FirstName, LastName, Middle Initial, School, Grade, Course Memberships, Entry Date, Exit Date, Course Enrollments (for Groups and
Rostering), LEA, Timestamp of last modification, Originator of modification

b) Parent Data from NC WISE: TBD – Still waiting to get all the input data before we can complete the specifications

c) Staff Accounts from UID: UID_ID, FirstName, LastName, Middle Initial, Department, Division, Position Site, Position Code, Position Description, Manager, Effective Start Date, LEA, School, Timestamp of last modification, Originator of modification

Of all the fields contained in each of these records, only those listed above might be consumed and used to populate the target systems.

FR2.4 Detect and handle changes in data between updates/consumptions (differentials)

The DI system must have the ability to determine deltas or changes between updates of user data. Assume the authoritative sources do not have the ability to distribute only changes in the user data.

Example: If the system consumes user data at 10 am and a single field/record changes with a student’s address, when the system consumes the user data again at 11 am the system must be able to detect that single changed record and update only the changed address.

FR2.5 Specifications

The DI system shall be able to determine deltas between successive inputs from authoritative data sources (e.g. NCWise, UID) that will likely consist of up to several million input data records per input. Input data set sizes will likely be on the order of up to 1-10 GB each, and may be in flat file format. Using the deltas that it computes, the DI system shall be able to propagate appropriate changes into and beyond the DI system.

FR5.5 Matching / Joining (Entity Resolution)

The DI system must match and join user data records as appropriate from disparate authoritative data sources. Typical data sources might include a Student System that also has Parent data, HR system for staff, Guest/Affiliate System, etc. The match or "join" process compares user data across these Systems of Record to ensure that a single record is created, modified, or deleted for an individual (e.g. in the Operational Data Store, IAM Person Registry, etc.), even if that individual exists in more than one System of Record simultaneously and/or moves back and forth between Systems of Record. The system should also be able to assign user affiliation (e.g. student, parent, employee, and guest) based on which system(s) of record the user is a part of or as a result of System of Record attributes. (Note that “affiliation” must be a multi-valued field to support concurrent affiliations).
Example: An individual is a parent in the Guest Registration system and that same individual is a Teacher in the HR system (additionally, parent information may be pulled from the student system). The DI System must provide a mechanism by which a single user record can be created, deleted, or modified for that individual, even though that individual exists in more than one System of Record.
FR3 Data Transformation Capabilities\textsuperscript{13}

Built-in capabilities for achieving data transformation operations of varying complexity, including:

- Basic transformations, such as data type conversations, strings manipulations and simple calculations
- Intermediate-complexity transformations, such as lookup and replace operations, aggregations, summarizations, deterministic matching and the management of slowing changing dimensions.
- Complex transformations, such as sophisticated parsing operations on free-form text and rich media.

In addition, the tools must provide facilities for developing custom transformations and extending packaged transformations.

\textit{Example:} The DI service would look into existing datasets, extract the elements and records defined by CEDS, convert that data into the defined CEDS formats and values, and transform those data into storable formats as defined by the CEDS XML schema, and deliver that data to the appropriate target service.

\textit{Example:} The DI service would look into existing datasets, extract the elements and records defined by Ed-Fi, convert that data into the defined Ed-Fi formats and values, and transform those data into storable formats as defined by the Ed-Fi XML schema, and deliver that data to the appropriate target service.

\textsuperscript{13} Gartner Magic Quadrant for Data Integration, October 2011, ID: G00219411
FR4 Load – Data Delivery Capabilities

The ability to provide data to consuming applications, processes, and databases in a variety of modes, including:

- Physical bulk data movement between data repositories
- Federated views formulated in memory
- Message-oriented movement via encapsulation
- Replication of data between homogeneous or heterogeneous database management systems (DBMSs) and schemas

In addition, support for the delivery of data across the range of latency requirements is important:

- Scheduled batch delivery
- Streaming / real-time delivery
- Event-driven delivery
FR5 Migration, Synchronization, and Replication

- The ability to migrate or synchronize between database structures, enterprise applications, mainframe legacy files, text, XML, message queues and a host of other sources.
- Metadata-driven access to sources and targets.
- Extensive library of predefined transformations can be extended and shared with other integration processes.
- Embedded, reusable data quality business rules clean data as it is moved, synchronized or replicated.
- Recognizes changes to key fields and replicates/synchronizes them across multiple databases.
- Integrated scheduler allows changes in one or more systems to be propagated to other systems on a scheduled basis.
- Delivers real-time data service for synchronization and migration projects.
FR6 Design and Development Environment Capabilities

Facilities for enabling the specification and construction of data integration processes, including:

- Graphical representation of repository objects, data models, and data flows.
- Workflow management for the development process, must address requirements such as approvals and promotions.
- Granular role-based and developer-based security.
- Team-based development capabilities, such as version control and collaboration.
- Functionality to support reuse across developers and projects, and to facilitate the identification of redundancies.
- Support for testing and debugging.
Developing a Data Integration Service

**FR7 Operational Data Store**

During the course of our research and planning for the DI system, we have determined that there is a need for a single, centralized, data repository in which consolidated data resides. Because a single data repository and data integration service for gathering user information does not exist, all target services are required to perform this additional functionality in order to create and manage data/information.

Each of the cloud services experiences this challenge. Data systems that handle HR and Finance will need to connect and retrieve user data for processing. Each data system would need to connect to multiple sources to consume the necessary data. Instructional data systems, such as an LMS or CMS, would need to connect and retrieve user data for processing. Each of these systems would also need to connect to multiple sources for getting the necessary data.

The same processes that were described in the previous functional requirements are repeated in the requirements throughout many cloud and other RttT systems. The process repetition creates several challenges as outlined below.

**Data Integration Duplication.** Without a single, centralized, data repository, each of the cloud and/or RttT systems will need to have a data integration component as part of its project. This integration tool must be complex enough to support the various disparate source systems. The cost and effort for each project is duplicated.

**Data Integration Interpretations.** Without a single, centralized, data repository, each of the cloud and/or RttT systems will need to consume the data and perform transformations of the data before provisioning into itself. There is a high probability that each of the systems will do this in a different way. As a result, two LMS systems could potentially have different user data for the same user.

**Security of Access.** Without a single, centralized, data repository, each of the cloud and/or RttT systems will need to have access to various source systems for user data. Therefore, each of these source systems would need to manage user authentication and authorization. In general, the DI system’s goal is to mitigate this situation. Managing the access for the different data sets is a challenging problem and could potentially lead to unauthorized access or data leakage. Having a single data repository would greatly simplify this process.
Alignment with National Shared Learning Infrastructure. On a national stage, the Gates and Carnegie foundations have partnered with the Council of Chief State School Officers (CCSSO) to support and fund a multi-state collaboration to develop and deploy a national level shared learning infrastructure (SLI). The CCSSO SLI work comprises content and data standards; content, data, and identity management infrastructure; a core set of “built-in” applications that leverage the content and data management infrastructure; and application interfaces supporting third-party development and integration with the SLI. Without a single, centralized data repository, the national SLI will have the same issues described above. A single, central data repository would enable closer alignment with SLI initiatives.

Recommendation: Design, develop, and deploy (or acquire as a service) a single, centralized data repository that contains an updated snapshot of operational data. The data will be fed from the various disparate source systems into the single, centralized repository. The repository will present reliable, contemporary, and standardized interfaces to the target systems for consuming the data.

FR7.1 Core Entity Data Model

A core entity data model must be developed/adopted for the Operational Data Store. Similar projects, such as the Shared Learning Infrastructure technology will use the Ed-Fi specification of CEDS for its core entity model. Another approach might be to use CEDS with SIF schema. This will need to be determined after more analysis of source and target systems has been completed.

FR7.2 API Requirements

The ODS shall include a set of application programming interfaces (APIs) to enable a secure and uniform interface for cloud services to easily access data. The ODS will support the following interfaces capabilities:

- RESTful Web service accessible over HTTPS
- Near-real-time, read-only access (Note: Any target that needs to send data back to ODS will be treated as a source system during initial stages of project)
- Asynchronous/batched access for bulk extracts
- School Interoperability Framework (SIF)
FR7.3 ODS and CEDARS (Data Warehouse)

The North Carolina Common Education Data Analysis & Reporting System (CEDARS) is NC’s PreK-13 State Longitudinal Data System. CEDARS is composed of various DPI source data collection systems, a student and staff identification system, a centralized data repository, and associated reporting and analysis (Business Intelligence) tools. Once completed, CEDARS will support NC’s efforts to use high quality data about students, staff, programs, and finances to make policy and service decisions that will improve student outcomes. Specifically, CEDARS will enable state, local, and federal policy makers and service providers to analyze trends and relationships between various educational factors and student performance over time.¹⁴

Developing a Data Integration Service

The challenge with CEDARS in relation to the Education Cloud projects is that it does not contain the most current version of some data. In general, CEDARS is concerned with longitudinal data and ODS is concerned with operational data. The ODS will complement CEDARS as a data repository that contains data currency from real-time in the source systems to where CEDARS currency begins. The ODS will fill that gap. The gap is different depending on the type of data being stored.

**ODS + CEDARS Gap Example**

![Diagram showing ODS and CEDARS gap example](image_url)
NFR1 Ed-Fi Data Standard\(^{15}\)

Ed-Fi is an XML-based universal data standard that facilitates data exchange among student data systems in the K-12 education sector. Development was funded by the Michael & Susan Dell Foundation, based on input from vendors, state education agencies and local education agencies. Ed-Fi augments existing systems instead of replacing them. Ed-Fi seeks to improve student achievement by extracting student information from a variety of sources, and by then standardizing, integrating and communicating this information to educators and other parties through Web-based dashboards, reports and other applications. Equipped with actionable insights, educators can make informed decisions on what steps are needed to improve the academic outcomes of students.

The Ed-Fi solution contains the following components:

- **Unifying data model (core schema)** – This is the central XML schema definition (XSD) that defines the common data elements such as the attributes, types and structures, on which the rest of the standard is based. The core schema serves the majority of schools’ educational needs and provides the building blocks for additional schemas, such as the interchange and extension schemas used for more specific situations.

- **Data exchange framework (interchange schemas)** – These are schemas that depend on the core schema and serve as connectors to send and receive information among systems. Interchange schemas differ from the core schema in that each interchange schema defines the data exchange in highly specific scenarios, while the core schema defines a broad set of objects used in many general situations. Vendors can create new interchange schemas to accommodate new scenarios.

- **Application framework (extension schemas)** – These are additional schemas that extend the XML types in the core schema and enable new types to be added. Extension schemas may be necessary to provide specific data for state education agency (SEA) reporting, or to address specific reporting needs for grants and other education programs.

- **Example relational database model with metrics and sample dashboards** – The Ed-Fi solution includes an example relational database model as well as the metrics needed to define performance management indicators. This toolkit provides the elements necessary to build dashboards.

**Relationship to DI:**

*The DI service would look into existing datasets, extract the elements and records defined by Ed-Fi, convert that data into the defined Ed-Fi formats and values, and transform those data into storable formats as defined by the Ed-Fi XML schema, and deliver that data to the appropriate target service.*

---

NFR2 Common Education Data Standard (CEDS)\textsuperscript{16}

The Common Education Data Standards (CEDS) project is a national collaborative effort to develop voluntary, common data standards for a key set of education data elements to streamline the exchange and comparison of data across institutions and sectors.

The Education Science Reform Act of 2002 gave the National Center for Education Statistics (NCES) the authority to determine voluntary standards and guidelines to assist state educational agencies in developing statewide longitudinal data systems (SLDSs). To this end, NCES is working with key stakeholders to develop standards for a core set of data elements to ensure that states create P-20 data systems that meet the goals of the American Recovery and Reinvestment Act of 2009. Standard data definitions will help ensure that data shared across institutions are consistent and comparable. This, in turn, will make it easier for states to learn how students fare as they move across institutions, state lines, and school levels.

**CEDS Alignment Tools**
https://ceds.ed.gov/alignmentTool.aspx

**Data Model Guide\textsuperscript{17}**
This guide includes a printable reference for the CEDS data model including the Domain Entity Schema (DES) and Normalized Data Schema (NDS).

\textit{Relationship to DI:}
\textit{The DI service would look into existing datasets, extract the elements and records defined by CEDS, convert that data into the defined CEDS formats and values, and transform those data into storable formats as defined by the CEDS XML schema, and deliver that data to the appropriate target service.}

\textsuperscript{16}https://ceds.ed.gov/whatIsCEDS.aspx
\textsuperscript{17}https://ceds.ed.gov/dataModel.aspx
NFR3 Schools Interoperability Framework (SIF)\(^{18}\)

The Schools Interoperability Framework Association is a non-profit organization that brings together vendors, government agencies, state departments of education, and other industry leaders to develop a specification that ensures that primary and secondary (K-12) instructional and administrative software applications can share information seamlessly. The SIF Association is not a product, but rather an industry-supported technical blueprint for primary and secondary (K-12) software that will enable diverse applications to interact and share data now and in the future.

The SIF Specification is a set of documents developed by SIFA work groups, committees and task forces comprised primarily of software engineers from educational software companies. These documents articulate a set of common definitions for school data and a set of rules for how this data can be shared. The common data definitions are called data objects. Data objects cover many items that are involved in schools. For example, a student’s name, address and phone number are part of the ‘StudentPersonal’ data object. Having different software programs understand this common definition of a student makes it possible for them to share this information properly. There are 89 data objects currently defined. Additional data objects will be defined as the Specification matures.

In addition to the data objects themselves, the SIF Specification also defines the rules for how software programs can send these data objects to each other. This set of rules is called the ‘infrastructure’ and uses ways of sending messages that are built on the types of technology utilized by the Internet. By using open and commonly available means to transport these data objects, SIF ensures that all vendors will be able to use the SIF framework and that all school systems will be able to implement it regardless of what kinds of computers or networks they have. Ensuring that SIFA is vendor-neutral and software-platform independent is an important guiding principal of SIFA and a foundation for the long-term viability of the SIF Specification.

SIF is a sophisticated, service-oriented architecture (SOA) technical standard for education. SIF ensures interoperability between disparate software applications — improving data accuracy and reliability, streamlining staff and business processes, and reducing costs.\(^{19}\)

**Relationship to DI:**

The ODS will support the SIF framework and enable the exchange of data between itself and target services. An assumption will be made that any target service that wishes to exchange data using SIF has the necessary components and configurations established.


\(^{19}\) [http://pearsondatasolutions.com/solutions/district-interoperability](http://pearsondatasolutions.com/solutions/district-interoperability)
NFR4 Scale and Scope

The K-12 user population comprises millions of users from NC’s 115 LEAs and 100+ charter schools, which include more than 2600 schools, 1.3 million students, 200 thousand faculty/staff, between 1-2 million parents/guardians, and additional stakeholders from the various local communities. The DI system must perform the data movements and processing to support this level of population. The DI system will initially support data integration for around 3 million users. It is expected that the user base would grow and churn to 10 million over the next 10 years.

Any platform providers must be able to provide evidence of previous successful DI deployments of this scale and scope. Any products/services will need to provide information related to large profile performance testing.

Any systems integrators or affiliated partners associated with the software solution must be able to demonstrate competency and evidence of a DI deployment of this scale and scope.

NFR4.1 Number of Sources and Targets
The DI software solution should be able to scale to any number of source and target systems with appropriate supporting infrastructure. There must be no restrictions inside the software solution that would limit the number of concurrent sources and targets on the system.
NFR5 Flexibility, Modularity, and Minimum Dependencies

The DI system must be designed and architected to be as flexible and modular as possible and demonstrate the concept of minimum dependencies. Minimum dependencies means, to the extent possible, the DI system will be designed in such a way that no single component is highly dependent on any other single component such that it could not be changed or modified with a reasonable amount of effort. The DI System will be designed in such a way that various components can be updated and/or changed with minimum impact on the overall system. It will have a plug and play feel. The design should allow the system to grow and adapt to new authoritative sources and targets services over time without requiring a re-architecture of the core system.

These same three concepts also apply to any system integrators, partners, or service providers. The systems and services surrounding the DI system will not be dependent or proprietary to any one organization or solution for support, service, or operations.

All aspects of the DI system will uphold the utmost continuity between the Cloud Services and the DI Service Provider.

*Example: We have Directory Service Software version 2011 and then version 2013 comes out. We must be able to update our software to the latest version with minimum impact to other components or services.*

*Example: Initial system consumes NCWISE data, but then NCWISE system is replaced with an alternate Student Information System (SIS). We must be able to update our software to consume data from the new SIS with minimal interruption or impact.*
NFR6 High Availability and Disaster Recovery

The DI system must be designed and implemented such that single component failures or planned maintenance events will result in minimal service disruptions. All scheduled maintenance activities will be planned, approved, and communicated well in advance. Also, any scheduled downtime must be aligned/coordinated within LEA and DPI maintenance windows. It should be noted that LEAs and DPI have different schedules and a maintenance schedule for one LEA or DPI might not work for another.

**NFRS6 High Availability Design Specifications**

While the design of the DI system shall be for 100% uptime, the system shall have a Service Level Agreement (SLA) specifying at least 99.9% uptime on a quarterly basis. Thus no more than 2.19 hours total per quarter of unexpected downtime would be acceptable according to the SLA. This may require development, testing, and redundant production environments.

NFR6.1 Platform Environments

The DI system must have a minimum of three DI system environments established. One environment would be used for testing, development, and general sandboxing. The second will be a pre-production/staging environment, configured in a highly available design to match the full production environment. The full production system will be the main environment that supports DI needs for Cloud services and targets.
NFR7 Change Management and Communication/Messaging

The DI system will have many different interactions with various source systems, target systems, user communities, and other affiliates. These interactions drive the need for solid communication/messaging and change management processes. The DI system must have comprehensive and efficient messaging and change management processes. All of the affected entities should be notified and made aware of changes within and associated with the DI System. DI Target Service Consumers should receive only information that is relevant to them, when needed and appropriate.

NFR8 DI System Performance

The DI system is expected to sustain a certain operational performance level. The backend components of the system are expected to consume and process data within certain time frames. At this time, the required information is not available to describe the specific performance parameters of the individual components or the system as a whole. However, the DI system and its components will be compared to each component’s/product’s documented performance profiles. The performance specification and metrics will be documented, as the necessary information is available.

NFR9 DI User Accessibility

The DI System user facing components must comply with relevant ADA regulations, Section 508 Amendment to the Rehabilitation Act of 1973, and align with the Web Accessibility Initiative (WAI) guidelines of the World Wide Web Consortium (W3C). The DI Systems must also support International Language for user facing components.

NFR10 Policy Support

Advisory bodies will provide oversight in terms of policies, rules and regulations surrounding the support, maintenance, and configuration of the NC Education Cloud Data Integration System. It will provide information for provisioning rules, workflows, processes, governance and data access issues for each of the major components of the system. The purpose of these bodies is to provide a governing structure for a sustainable system. The DI system must be flexible enough to adapt to the advisory body’s requirements as they are developed and adapted over time.
PART III

SERVICE MODEL
Developing a Data Integration Service

**DI Service Model Overview**

Through a substantial research and planning effort, we have a good understanding of the data integration challenges in K-12 based educational systems and know what the design and architecture of a DI system should contain (DI Project – Phase I). Now we need to describe a model for supporting and managing the system/service in a sustainable manner that continuously brings value to the stakeholders. We begin by describing the strategies used when forming the DI Service Model (DI Project – Phase II). Next, it is important to understand the participants in the Service Model. We discuss each of the service roles and their characteristics.

**Strategic Principles**

Throughout the service model section we will be describing the different aspects of supporting and managing a DI service. There are some common principles applicable to each.

**Sustainability.** It is very important that the DI system/service be designed, built, deployed, and managed in such a way that is functionally and economically sustainable for many years to come. Earlier in the DI architecture plan, we discussed details around minimal dependencies, modular design, and redundant systems. This theme will carry over into the Service Model. Looking at each component of the service model individually but also holistically (even outside of the DI project), will allow us to create a sustainable system and service.

**Value.** In addition to being sustainable, is important that the central system and service continuously bring value to the stakeholders and consumers.

**Purpose.** All planning for this project ultimately connects back to the purpose/vision of the DI Service and the NC Education Cloud. From the beginning, the project has set out to solve three challenges faced by the NC Education community: reliable and standard data availability and access, limited resources for performing data integration, and having a solid foundation for other cloud services. The Service Model will continue to support efforts to address those challenges.

**Community.** A system of this scale and scope requires an enormous amount of planning to be successful. Most of the planning work includes countless interactions with the various organizations and stakeholders that will ultimately play a role in the DI Service. Getting the community involved in the planning process has increased the accuracy of requirements/specifications and created the best alignment with business processes while bringing new value to the community. This also confers ownership of the DI system/service to the communities that it serves.
Policy and Governance. There are many different components and organizations that contribute to the DI system/service. It is important that a well-defined policy and governance model be in place that clearly defines how those components and organizations work together to effectively deliver a sustainable and valuable service.

Measures of Success. Each step of the Build and Implementation will clearly define measurable points, priorities, and acceptance criteria. By defining these at each step, we have success indicators and criteria to know when each step has been completed.

Scope. The DI Service has a well-defined scope, is achievable, and promotes value to the stakeholders. The project scope lays a good foundation to continuously build upon with new and additional features around data integration.
DI Service Roles

The DI system will be implemented and deployed as a service to the NC Education Cloud Services, among other consumers. The DI service is composed of five service roles. Each “role” represents a collection of organizations (or components, systems, etc. - owned by organizations), that will serve a function or purpose in the DI system/service. These organizations will work collaboratively but have distinct responsibilities in building, deploying, and managing the DI Service.
Developing a Data Integration Service

DI Policy and Governance Role

DI Policy and Governance is a collection of heterogeneous, cross-functional advisory bodies that oversee, support, and “create and endorse policies”, to ensure an effective and efficient DI system and service. The current bodies are the DI Working Group, NC Education Cloud Team, NC Education Cloud Advisory Committee, NC eLearning Commission, Governors Education Transformation Committee, and NC State Board of Education. In the future, these advisory bodies may change as the NC Education Cloud project ends and the responsibilities/oversight transitions to an Education Service Agency.

Service Roles Partners: DI Data Sources, DI Service Manager, DI Target Services

The DI Policy and Governance bodies will be provided:

- Periodic reports about the DI Service by the DI Service Manager describing service progress, issues, successes, funding status, future plans, and recommendations
- Federal or State required reports
- Advice on what policy decisions are needed for the delivery of the DI Service
- General feedback from the DI Target Service about the quality of the DI Service provided, including support
- Advice on what policy decisions are needed from the DI Target Services

The DI Policy and Governance will be responsible for:

- Developing, updating, and maintaining a Policy and Governance structure for the DI Service that supports the needs of the DI Target Services
- Developing and defining policies and procedures needed for the DI Service
- Defining and refining the DI Service Roles
- Providing requirements to DI Service Manager for the DI Service
- Providing input and review of any DI Managed Service Contracts
- Serving as a configuration control board for the DI Service components
- Serving as a DI Vision Group for shared decision-making and priority setting
As necessary, diplomatically working with different personalities among constituents to achieve end goals and keep the project/service moving forward.

As mentioned earlier, each of the service role organizations will work collaboratively but have distinct responsibilities. The chart below shows the authoritative relationships between the services roles, in the context of the DI system.

The DI Policy and Governance role members are the ultimate authority bodies that give direction to the DI Data Sources, DI Service Manager, and DI Target Services. This direction will come in different forms depending on the originating body and nature of the communication. It is important to note that each of the service roles will have representatives in Policy and Governance. The DI Managed Service will take its direction from the DI Service Manager as granted by the DI Policy and Governance bodies.
**DI Service Manager Role**

The DI Service Manager is an organization that serves as an intermediary between all DI Service roles. The current market offerings for DI Systems do not align well with the DI Target Service needs and the DI Service Manager will serve to bridge any gaps that exist, to enable a more complete, sustainable and valuable solution.

*Service Roles Partners: DI Policy and Governance, DI Data Sources, DI Managed Service, DI Target Services*

The DI Service Manager will have the following *qualifications*:

- Education Community Focused
- Previous experience delivering state-wide services as a K-12 Service Provider/Manager
- Excellent history of customer service in the K-12 community
- Understand K-12 Users and Business Processes and Business Cycles

The DI Service Manager will be *provided*:

- Requirements from DI Policy and Governance (and DI Target Services when appropriate)
- Policy rules/decisions/endorsements from DI Policy and Governance
- A Technical Implementation Plan created by the DI Managed Service provider (in conjunction with the DI Service Manager) that details the individual steps to be performed by the Managed Service provider in installing and configuring the DI System/Service
- Documentation and knowledge transfer from the DI Managed Service Provider
- Payment/Funding for the delivery of the DI Service

The DI Service Manager will be *responsible for*:

- Managing an NC Education DI Service as specified by the NC Education Cloud and other advisory bodies
  - Pilot then Production – phased approach
Developing a Data Integration Service

- Overall management of the DI Managed Services
- Establishing relationships and coordinate as needed with owners of authoritative source data systems
- Providing technical support services and a help desk (may be part of a larger NC Education Cloud Help Desk), to the DI Target Services for matters related to the DI Service
- Ensuring the implementation and support of policies as defined by DI Policy and Governance bodies
- Providing technical and strategic guidance to the DI Policy and Governance bodies about any necessary policies that are needed to help the DI Service Manager deliver and support a better DI Service for the DI Target Services
- Coordinating with DI Managed Service provider to translate DI Target Service and/or DI Policy and Governance requirements into enhancements or corrections for the DI Service
- Coordinating with DI Managed Service provider and DI Target Services to onboard/integrate new Cloud Applications, Resources, and Services
- Providing to DI Target Services and DI Policy and Governance bodies any reporting and audit information as needed (may be provided by a Cloud Security and Compliance group)
- Providing direction and documentation to DI Target Services as they onboard and use the resources/services provided by the DI Service
- Tracking suggestions for improvements, features, and developing the methodology for DI system enhancements
- Developing metrics and reports on the impact of DI on the Target Services to DI Policy and Governance where appropriate
- Communicating new DI Service features to appropriate DI Target Service participants
- Managing vendor churn for DI Service components
- Managing vendor contracts and developing acceptance criteria for each checkpoint of a contract
**DI Data Sources Role**

The DI Data Sources are the authoritative systems that contain K-12 user data to be processed by the DI system. The NC Department of Public Instruction (DPI) currently hosts and maintains the centrally located source systems. The DI Service Manager will coordinate with DPI and the DI Managed Service provider to transfer required user data from the source systems to the DI system.

*Service Roles Partners: DI Policy and Governance, DI Service Manager, DI Managed Service, DI Target Services*

The DI Data Sources will have the following qualifications:

- Standardized data integration interfaces and methods
- Products/services with documented APIs for data extraction

The DI Data Sources will be provided:

- Notice/communication about DI Service Maintenance
- Appropriate requests for necessary data sets and fields needed in accordance with data source policies
- Feedback regarding data source feeds including any data irregularities

The DI Data Sources will be responsible for:

- Supplying required user data sets and attributes at the specified frequencies to the DI System
- Supplying support for user data issues and authoritative source systems
- Change Management notifications for any DI-connected source systems, data structures, or formatting standards
- Notifications of changes of “authorized person” to receive notifications to avoid blackhole syndrome
DI Managed Service Role

The DI Managed Service (Provider) is the organization(s) that provide the necessary technology components that compose the DI service and technical support for managing the components. These could be software vendors, technical product architects, systems integrators, and/or infrastructure hosting providers. All these components together will be presented as a single service.

Service Roles Partners: DI Policy and Governance (indirectly), DI Data Sources, DI Service Manager, DI Target Services

The DI Managed Service Provider will have the following qualifications:

- Access to internal vendor support and development groups
- *Common Criteria Certified* for DI Software Components

The DI Managed Service will be provided:

- Payment/Funding for DI Components and Services provided to the DI Target Services and the DI Service Manager
- Requirements and Specifications from the DI Service Manager
- Expectations, Rules, and Policy surrounding the security and protection of the DI System’s data information

The DI Managed Service will be responsible for:

- Providing to the DI Service Manager and DI Target Systems the DI System as a Service
- Providing the necessary cloud-hosted infrastructure, along with technical support for services provided and the service, maintenance, and operational support for the software
- Performing the initial design and implementation of the DI system as specified by the DI Service Manager and DI Policy and Governance bodies and as described in a Technical Implementation Plan *document* to be delivered to the DI Service Manager prior to any actual work beginning
• Providing the initial DI application administration and configuration, with the possible transfer of this responsibility over time, to the DI Service Manager, as the system matures

• Knowledge Transfer and Documentation about DI Service components to the DI Service Manager

• Providing recommendations on the numbers and skillsets of staff to support the DI Service through implementation and ongoing production

• Performing data profiling and transformations of source system data extracts

• Performing target system application integration to the DI System
DI Target Services Role

The DI Target Services are any application or cloud service that connect to the DI system and/or utilizes the DI service for data transformation or integration needs. Other state agencies may have staff members that utilize the DI system to access education data resources.

Service Roles Partners: DI Managed Service, DI Data Sources, DI Service Manager, DI Policy and Governance

The DI Target Services will be provided:

- Technical support and documentation that illustrate how each data entity interacts with its respective area of the DI system
- Requirements for integrating Cloud Services with the DI System
- Policy and Governance requirements related to the DI Service and DI Target Services
Developing a Data Integration Service

The DI Target Services will be *responsible* for:

- Assurance that user data is managed in the most secure and compliant manner
- Adhering and abiding to the rules set by the DI Policy and Governance bodies regardless of type of cloud service, and any contract requirements regarding the use of user data
- Providing notification and messaging to the DI System and Service Manager about cloud service general changes, maintenance, and outages
- Providing a dependable, responsive and available service (application), to ensure a successful and consistent user experience

Over time as the DI project matures and the actual needs are better known, there will be an evolutionary shift in workload and responsibilities among the service roles.

Now that we understand who the service role organizations are, their expectations, and general responsibilities, we are ready to examine the other major tasks of the service model.
Developing a Data Integration Service

Project Portfolio Management

The DI team will be coordinating with the State EPMO throughout the lifecycle of the DI project for project management requirements and procurement needs.

“The Enterprise Project Management Office (EPMO) was established in 2004 to assist the State CIO (SCIO) in his legislated responsibility to improve the management of IT projects in state government. The EPMO is one component of this legislation; NC General Statutes - Chapter 147 Article 3D.

The State CIO has legislative oversight authority to review and approve State agency IT projects; develop standards and accountability measures for IT projects (including criteria for adequate project management); to require status reporting; to assign a project management advisor (PMA); and, if necessary, to suspend projects. The purpose of this legislation is to help ensure that quality IT projects are delivered in a cost effective and timely manner.

By law, agencies must provide a project manager who meets the applicable quality assurance standards for each IT project. The agency project manager is responsible for managing the project and providing periodic reports which shall include information regarding project costs; issues related to hardware, software or training; projected and actual completion dates and any other information related to the implementation of the IT project.

The EPMO provides professional oversight to facilitate successful IT projects through assessing projects and facilitating resolution of issues, risks and roadblocks and by providing early warnings if a project is likely to fail. The EPMO does this through applying professional experience and best practices and by providing full and complete information to the SCIO to ensure the SCIO can respond appropriately.

The EPMO also engages agencies in understanding and applying IT portfolio management concepts to IT projects, applications and IT investments (plans and budgets). This is done through IT Portfolio Management services, which provides financial, technical, and management personnel with a knowledge and understanding of the theories, concepts, and disciplines of portfolio management. The service also includes training and support for the use of the Portfolio Management software tool to improve the planning, budgeting, and management of IT, especially IT investments.”

20 http://www.epmo.scio.nc.gov/
State of North Carolina PPM Workflow for Projects >= $500,000

**Phase 1: Project Initiation**
- Gate 1: Planning & Design Approval
  - SCIO, OSBM, OSC, EPMO, EA

**Phase 2: Planning & Design**
- Gate 2: Execution & Build Approval
  - SCIO, OSBM, OSC, EPMO, EA

**Phase 3: Execution & Build**
- Gate 3: Implementation Approval
  - SCIO, OSBM, OSC, EPMO, EA

**Phase 4: Implementation**
- Closeout Approval
  - EPMO QA Signoff

**Phase 5: Closeout**
- Closeout Signoff
  - Agency Contributors

**Initiation Exit Criteria and Planning & Build Exit Criteria**
- Project Information Tab
  - Project Charter (Business Issues, Business Goals, Project Goals, Project Deliverables, Items Out of Scope, Proposed Strategy, High Level Assumptions / Constraints, Key Dependencies external to the project, Project Organization and Roles)
  - Initiation Phase Budget
  - Planning & Design Phase Budget
  - Budget Information Section
  - Enterprise Architecture Questionnaire
  - Security Questions
- Status Tab
  - Project/Phase Milestones & Key Deliverables
  - Plan vs. Actual
- Issues & Risks Tab
  - Update any issues and risks that are open
  - Strategic Impact Tab & Risk Tab
  - Open any issues and risks that may have changed
  - Document Management Tab
    - Updated Staffing Plan
    - Updated Procurement Plan
    - Approved Business Requirements Document
    - Technical Architecture System Design Doc
    - Sponsor’s Approval of User Acceptance Criteria
    - Documents checked in the Project Info tab or identify where the document is stored.
- Budget, Scope and Schedule are Baseline at Gate 2

**Agency Document Checklist**
- Disaster Recovery/Business Continuity Plan
- Acceptance Criteria
- Quality Assurance Plan

**AGENCY DOCUMENT CHECKLIST**
- Project Management Plan
- Communication Plan
- Change Mgmt Plan
- Configuration Mgmt
- Project Schedule
- Acceptance Criteria
- Quality Assurance Plan

**Implementation Exit Criteria & Closeout Entrance Criteria**
- Project Information Tab
  - Updated Project Charter
  - Updated Budget Information Section
  - Updated Enterprise Architecture Questionnaire
  - Updated Security Questions
  - Financial Benefits must be entered at Level 2
- Status Tab
  - Project Schedule / Milestones & Key Deliverables
  - Monthly Status Reports Current Cost Tracking Tab & Cost Forecast Tab
  - Plan vs. Actual
  - Issues & Risks Tab
  - Open any issues and risks that may have changed
  - Document Management Tab
    - Updated Staffing Plan
    - Updated Procurement Plan
    - Updated Technical Architecture System Design Doc
    - Sponsor’s Approval of O&M Q&A
  - Final project costs need to be entered
  - Load historical documents in the Document Management Tab

**NOTES**
1. **Template provided in the PPM Tool: NC HELP section and on EPMO website: www.epmo.scio.nc.gov** is required for these documents.
2. All projects >= $500K are required to submit monthly status reports.

**Version 4.4 06/08/2011**
Service Plan Development

The DI Service Manager will develop a Service Plan that describes the details of how – with the DI Managed Service provider - it plans to develop, deploy, and manage the DI system/service for the DI Target Services. As previously mentioned, some items in the Service Plan are dependent on details of the Technical Implementation Plan. The Service Model part of this DI plan will serve as the initial draft and framework for the more detailed Service Plan.

Requirements of Service Plan

Below are expected requirements that should be outlined in the DI Service Manager Service Plan.

Communications Model. Messaging, notifications, and alerts are all communication components of the DI Service. The service plan will need to describe how the DI Service Manager plans to effectively and accurately communicate information inside its own organization as well as to and from Service Roles for all aspects of the DI system. Open feedback mechanisms will be important tools for continuously improving the DI Service over time.

Support Model. While DI Policy and Governance will establish many of the rules for support, the service plan needs to describe the processes and models for supporting the other Service Roles of the DI System.

Outreach, Direction and Documentation. Closely related to support, the DI Service Manager will need to provide direction and documentation related to the different Service Roles. The service plan should describe the expected types of training and documentation it will provide along with other outreach and knowledge management activities/efforts.

DI Service Pilots. The DI system must be piloted and vetted with DI Target Service Cloud Services (applications) before moving to full production mode. The service plan must describe details of the different DI Service Pilots and how they relate/contribute to the overall DI System/Service.

Milestones and Deliverables. The service plan must describe the different milestones and deliverables for each phase of the implementation, deployment, and management of the DI Service.

Personnel Requirements. The service plan must describe the internal personnel requirements to effectively manage the DI Service initially and long-term. The personnel requirements should include each person’s role and responsibilities within the DI Service along with the necessary skills set needed.
**Sustainability and Value.** The service plan must define strategies of the DI Service Manager that will enable it to provide a long-term sustainable and continuously valuable DI Service to the DI Target Services.

**Funding Model.** The service plan must include a detailed budget that reflects how the DI Service Manager intends to use the project funds to deliver the service functions requested.

**Biographical Sketch.** The service plan must provide a biographical sketch and qualifications of the DI Service Managers current staff that would be involved in providing the DI Service.

**Commitments.** The service plan must provide a description of other commitments by the DI Service Manager as a metric of availability to the DI Service.

**Change Management.** The service plan must describe the DI Service Manager’s change management policies for support systems of the DI Service.

**Software Asset Management.** The service plan must describe how the DI Service Manager intends to manage and optimize the purchase, deployment, maintenance, utilization, and disposal of software applications with the DI System/Service.

**Acceptance Testing and Metrics.** The service plan must describe how the DI Service Manager plans to develop metrics for acceptance testing of products and services from the DI Managed Service.
Data Sources and Target Cloud Services

One major task of the DI Service Model phase will be to formally establish agreements with authoritative data providers and initial targeted cloud services. This task will allow the DI team and Managed Service Provider to clearly define requirements and scope for the Technical Implementation Plan and Phase III of the DI project - Building and Implementation.
Data Sources

Data Sources are the K-12 authoritative sources where user data and attributes reside. The following systems will be the main authoritative data sources for DI:

**NCWISE**
http://www.ncwise.org/

First introduced to North Carolina Public Schools in 2004, the North Carolina Window of Information on Student Education (NC WISE) integrates all aspects of public school life from the classroom to the central office. Web-based and centrally maintained for capturing, accessing, and reporting a wide spectrum of student information, NC WISE replaces the Student Information Management System (SIMS), which North Carolina Public School Systems have used for almost two decades.

In 2009, NC WISE completed a statewide rollout, which includes all of the state’s 115 LEAs and 98 charter schools.

NC WISE is composed of three basic components:

- Electronic Student Information System (eSIS) allows individual schools to manage student information.
- Electronic Data Interchange provides the capability to electronically transmit all student instructional records and demographic information between school districts, between schools within a district, and to universities and colleges in North Carolina.
- Uniform Education Reporting System (UERS) transfers information from the local school district to the North Carolina Department of Public Instruction (NCDPI).

Essentially, NC WISE integrates all aspects of public school life from the classroom to the central office.
The NC Department of Public Instruction (DPI) has developed a Unique Statewide Student and Staff Identifier (UID) System that will be the cornerstone of the NC Common Education Data Analysis & Reporting System (CEDARS). The UID System enables LEAs and Charter Schools to assign a unique statewide identifier for every student and staff member who participates in the NC education system at any point in time. Establishing the capability to assign these identifiers is the first step in DPI’s multi-stage effort to create CEDARS. The UID enables reliable matching of student-level and staff-level records over time and across DPI applications. This capability is improving data quality at the state and local levels and supports state and local policy makers and service providers in making better data-driven decisions.

NC Staff Identification System (Staff UID)

- The Unique Identifier for Staff System (Staff UID) will assign a unique identifier to staff who participate in the North Carolina public school system.
- UIDs follow staff between school districts and remain valid even if they move out of state and then return to a NC public school.
- Staff UID is built on the eScholar Uniq-ID® for Staff product.
CECAS
http://www.nccecas.org/geninfo.html

The Comprehensive Exceptional Children Accountability System (CECAS) is a case management and data analysis system that will be offered to Local Education Authorities (LEAs), Charter Schools, and State-Operated Programs (SOPs) as a means to manage and analyze exceptional children data. This project customizes a commercial product to provide efficient tracking of services for exceptional children. The application is a completely outsourced, web-based system.

EDDIE
http://www.ncpublicschools.org/fbs/accounting/eddie/

The Educational Directory and Demographical Information Exchange (EDDIE) is an online application containing LEA (school district) and school information such as LEA numbers, school numbers, administrative contacts, school types, grade levels, calendar types, program types, addresses, and more. This information is maintained and edited by subscribed users at the LEAs and Charter Schools.

EDDIE is the authoritative source for NC public school information and feeds multiple systems including: Accountability, NCWISE, NC School Report Cards, and is used to meet federal reporting requirements.

ACC
http://www.ncpublicschools.org/accountability/

TBD
Shared Learning Infrastructure
http://www.slcedu.org

The Shared Learning Collaborative (SLC) is an alliance that aims to accelerate the progress of our public schools toward personalized learning for all students. The collaborative is designing a shared technology infrastructure that will support the implementation of the Common Core State Standards and help states and districts provide teachers with the instructional data and tools they need. Instructional data will be linked to high-quality and diverse sets of curricular resources, so each student gets what he or she needs most at that moment in time. And, because the Common Core State Standards create more consistency across states than ever achieved in the past, we have an opportunity to create an infrastructure that works better and costs less per state than what can be accomplished by each state working individually.

The SLC is building a set of shared services that will connect disparate student data and learning content that currently exist in different formats and locations and don’t integrate with one another. The SLC technology will include the following:

**Middleware:** Software that integrates and orchestrates activities across different state systems, components and applications, enabling them to interact

**Data store:** A secure, cloud-based repository for structured and unstructured learning data

**Dashboards:** Out-of-the-box dashboards to make student data more manageable and useful for educators in a customizable format

**Learning maps:** Graphical representations of student learning data to help visualize student achievements and needs

**API:** An open API to enable vendors and developers to create applications and content that can interface with the SLC technology

The initiative is a state-led effort. The Council of Chief State School Officers is coordinating the multistate alliance and ensuring the new technology meets the needs of the states and reflects their input. Initially, Colorado, Delaware, Georgia, Illinois, Kentucky, Louisiana, Massachusetts, New York and North Carolina are participating in the pilot of this collaborative effort that will eventually be available to all states. The Bill & Melinda Gates Foundation and Carnegie Corporation of New York are providing the initial funding.
IIS (current NC RttT Project)
http://www.ncpublicschools.org/acre/improvement/

The North Carolina Instructional Improvement Systems (IIS) will be:
An online resource for students, parents, teachers, and administrators and will house instructional and learning resources, technology-based tools, and other strategies to provide meaningful support and actionable data to make informed decisions related to improving instruction and learning. Such activities include:

- Instructional planning
- Gathering information using a balanced assessment approach [e.g., through formative assessment (as outlined in NC FALCON), interim assessments (utilizing item banks that will be made available to all districts), summative assessments (at the classroom and state levels)]
- Displaying and analyzing information through user friendly dashboards using these data to inform decisions on appropriate next instructional or learning steps and career and college readiness
- Evaluating the effectiveness of the actions taken.

The NC IIS will promote collaborative problem solving and action planning. The system will also integrate instructional and learning data with student-level data such as attendance, discipline, grades, credit accumulation, and student survey results to provide early warning indicators of a student’s risk of educational failure.
CEDARS
http://www.ncpublicschools.org/cedars/

The Common Education Data Analysis and Reporting System, known as CEDARS, is North Carolina's PreK-13 State Longitudinal Data System. The system is composed of various DPI source data collection systems, a student and staff identification system, a centralized data repository, and associated reporting and analysis (or "business intelligence") tools.

CEDARS supports NC’s efforts to use high quality data about students, staff, programs, and finances to make policy and service decisions that will improve student outcomes. Specifically, CEDARS will enable state, local, and federal policy makers and service providers to analyze trends and relationships between various educational factors and student performance over time.

Establishing the Unique Identifiers (UID) for student and staff was the first step in the NC Department of Public Instruction's (DPI) multi-stage effort to create CEDARS. These identifiers will enable matching of student and staff data across DPI data collections and over time. The CEDARS Data Warehouse and Reporting System contains various DPI source data in a centralized data repository. CEDARS will, over time, increase local capacity to track student performance and assist local educational leaders in making data-driven decisions.
AMTR 3.0 (current NC Education Cloud Project)
http://cloud.fi.ncsu.edu/projects/amtr.php

The Data Collection and Analysis Project is purposed with revamping various technology oriented survey and planning tools. This work builds on the long standing Annual Media and Technology Report (AMTR) as well as recent survey work done in conjunction with the North Carolina Education Cloud Program.

This project has three major goals: 1) Overhaul the AMTR to be more “cloud-centric” and integrate the total cost of ownership tools into the survey; 2) Develop a business process tool that integrates LEA Technology Plans, survey data, and financial data; 3) Provide an interface to these datasets in an easy to understand format, and that provides views relevant to school, district, and state level audiences.

Development of the business process tool will require further planning and work with LEAs to develop. The addition of relevant datasets, such as attendance, e-Rate, finance, etc., will be ongoing work as well.
Developing a Data Integration Service

The list above describes each of the confirmed source data systems to be integrated with the DI Service initially. However, there are many other applications, systems, and cloud services that could be potentially integrated as well. Below is a list of some current applications/services that DPI is using that could be part of the next potential set of application services integrated with DI.

- SNA - School Name and Address system manages personnel contacts and addresses for LEAs and schools.
- LIC/SAL - Licensure and Salary system contains teacher pay history and licensure information.
- MIS2000 - Collection of information for children identified as migrant.
- CTE - Career and Technical Education data pertains to enrollments and performance of CTE students.
- Child Nutrition/Free and Reduced Lunch - Collection of data that shows students who are participating in the Free and Reduced lunch program. This information is used to determine economically disadvantaged status.
- Homeless - Collection of information for students identified as homeless.
- Immigrant - Collection of information for students identified as immigrants.
- LEP - Collection of information for students identified as limited English proficient.
- Neglected and Delinquent - Collection of information for students identified as neglected and delinquent.
- SES/Choice - Information on Supplemental Education Services and School Choice offered to students.
- Title I/TAS - A collection of schools in either a school-wide Program or Targeted Assistance Program, which captures those schools that have students specifically targeted by Title I.
- Dropout - Supports the annual dropout data collection as required for state and federal dropout reporting.
- DBS Internal and External - Financial data submitted monthly by the LEAs and Charters.
- PPAR - The Professional Personnel Activity Report is a collection of data for schools that cannot submit their data due to scheduling/track issues or other reasons, through the School Activity Report.
- PMR Data - Verified data from NCWISE about enrollment, membership, and attendance by grade level from each Public School.
- MFR/AFR - Financial data submitted monthly by the LEAs and Charters.
Target Cloud Services

Target Cloud Services are the web-based applications, systems, and services that will be utilizing the Operational Data Store. We expect initial cloud services will likely include some or all of the following, in addition to applications to be determined in coming months that are not yet included in this table.

Moodle
http://moodle.org

Moodle is a Course Management System (CMS), also known as a Learning Management System (LMS) or a Virtual Learning Environment (VLE). It is a free web application that educators can use to create effective online learning sites.

http://www.moodlerooms.com/

Moodlerooms is an organization that combines tools, expertise, a proven process, and a partner network to create a solution to help expand, support, and share the world's most widely used online learning platform, Moodle.

Blackboard
http://www.blackboard.com/

Blackboard Inc. is a global leader in enterprise technology and innovative solutions that improve the experience of millions of students and learners around the world every day. Blackboard's solutions allow thousands of higher education, K-12, professional, corporate, and government organizations to extend teaching and learning online, facilitate campus commerce and security, and communicate more effectively with their communities.

ClassScape Assessment System
http://www.classscape.org

ClassScape is an online assessment system that facilitates learning by focusing on curricular objectives. ClassScape enables teachers to monitor individual student and class performance on the North Carolina Standard Course of Study (NCSCS) goals and objectives. North Carolina teachers can build their own customized assessments or use ClassScape’s prepackaged tests to assess student performance on specific objectives at any point during the year.
Developing a Data Integration Service

**SchoolDude**
http://www.schooldude.com/

SchoolDude.com is a provider of online tools for management of school operations. SchoolDude’s entire suite supports a wide range of software that streamlines every aspect of a school’s functional areas – including enterprise-wide technology, facilities, energy, and event management support services.

**SAS EVAAS**
http://www.sas.com/govedu/edu/k12/evaas/index.html

As the most comprehensive reporting package of value-added metrics available in the educational market, SAS EVAAS for K-12 provides valuable diagnostic information about past practices and reports on students’ predicted success probabilities at numerous academic milestones. By identifying which students are at risk, educators can be proactive, making sound instructional choices and using their resources more strategically to ensure that every student has the chance to succeed.

**SAS Curriculum Pathways**

SAS delivers online curriculum resources with a proven impact on increasing student learning and teacher effectiveness – at no cost to US middle and high schools, virtual schools, home schools and other teaching and learning environments. SAS Curriculum Pathways is designed to enhance student achievement and teacher effectiveness by providing Web-based curriculum resources in all the core disciplines, to educators and students in grades 6-12.

**IIS (current NC RttT Project)**
http://www.ncpublicschools.org/acre/improvement/

The Instructional Improvement System (IIS) is one of Governor Perdue's Career and College: Ready, Set, Go! projects being developed through North Carolina's Race to the Top grant award. The IIS will provide portals for students, teachers, parents, and school and district administrators to access data and resources to help with decision-making related to instruction, assessment, and career and college goals.
IAM (current NC Education Cloud Project)
http://cloud.fi.ncsu.edu/projects/iam.php

The NC Education Cloud Identity and Access Management System (IAM) shall provide every K-12 student, teacher, staff member, parent/guardian, and school community member in North Carolina an account, with a single username and password, that will enable access to cloud-based learning resources. The IAM system will have three major components: a centralized data repository with all user information, a central directory service that provides a master authentication and authorization resource, and federation software that enables Single Sign-On functionality for users.

AMTR 3.0 (current NC Education Cloud Project)
http://cloud.fi.ncsu.edu/projects/amtr.php

The Data Collection and Analysis Project is purposed with revamping various technology oriented survey and planning tools. This work builds on the long standing Annual Media and Technology Report (AMTR) as well as recent survey work done in conjunction with the North Carolina Education Cloud Program.

This project has three major goals: 1) Overhaul the AMTR to be more “cloud-centric” and integrate the total cost of ownership tools into the survey; 2) Develop a business process tool that integrates LEA Technology Plans, survey data, and financial data; 3) Provide an interface to these datasets in an easy to understand format, and that provides views relevant to school, district, and state level audiences.

Development of the business process tool will require further planning and work with LEAs to develop. The addition of relevant datasets, such as attendance, e-Rate, finance, etc. will be ongoing work as well.

CEDARS
http://www.ncpublicschools.org/cedars/

The Common Education Data Analysis and Reporting System, known as CEDARS, is North Carolina's PreK-13 State Longitudinal Data System. The system is composed of various DPI source data collection systems, a student and staff identification system, a centralized data repository, and associated reporting and analysis (or "business intelligence") tools.
Shared Learning Infrastructure
http://www.slcedu.org

The Shared Learning Collaborative (SLC) is an alliance that aims to accelerate the progress of our public schools toward personalized learning for all students. The collaborative is designing a shared technology infrastructure that will support the implementation of the Common Core State Standards and help states and districts provide teachers with the instructional data and tools they need. Instructional data will be linked to high-quality and diverse sets of curricular resources, so each student gets what he or she needs most at that moment in time. And, because the Common Core State Standards create more consistency across states than ever achieved in the past, we have an opportunity to create an infrastructure that works better and costs less per state than what can be accomplished by each state working individually.

The SLC is building a set of shared services that will connect disparate student data and learning content that currently exist in different formats and locations and don’t integrate with one another. The SLC technology will include the following:

Middleware: Software that integrates and orchestrates activities across different state systems, components and applications, enabling them to interact

Data store: A secure, cloud-based repository for structured and unstructured learning data

Dashboards: Out-of-the-box dashboards to make student data more manageable and useful for educators in a customizable format

Learning maps: Graphical representations of student learning data to help visualize student achievements and needs

API: An open API to enable vendors and developers to create applications and content that can interface with the SLC technology

The initiative is a state-led effort. The Council of Chief State School Officers is coordinating the multistate alliance and ensuring the new technology meets the needs of the states and reflects their input. Initially, Colorado, Delaware, Georgia, Illinois, Kentucky, Louisiana, Massachusetts, New York and North Carolina are participating in the pilot of this collaborative effort that will eventually be available to all states. The Bill & Melinda Gates Foundation and Carnegie Corporation of New York are providing the initial funding.
Developing a Data Integration Service

**ODS** (current NC Education Cloud Project)

TBD per DI Project

Each of the Target services described above represents an application that is delivered as a cloud service offering. This means the DI system will integrate with any of the above applications that LEAs are purchasing as a cloud service offering or offered by DPI.
The list above describes each of the confirmed target cloud services to be integrated with the DI Service initially. However, there are many other applications/cloud services that could be potentially integrated as well. Below is a list of some current applications/services that the LEA community is using and could be part of the next potential set of application services integrated with DI.

<table>
<thead>
<tr>
<th>WikiSpaces</th>
<th>Skype</th>
<th>SharpSchool</th>
<th>OdysseyWare</th>
</tr>
</thead>
<tbody>
<tr>
<td>SchoolWires</td>
<td>PBwik/PBworks</td>
<td>Tandberg</td>
<td>Classworks</td>
</tr>
<tr>
<td>Edmodo</td>
<td>WebEx</td>
<td>Blogspot</td>
<td>Worldbook</td>
</tr>
<tr>
<td>SchoolCenter</td>
<td>GoToMeeting</td>
<td>Polycom</td>
<td>Atomic Learning</td>
</tr>
<tr>
<td>SharePoint</td>
<td>PBwik/PBworks</td>
<td>Edublog</td>
<td>BrainPop</td>
</tr>
<tr>
<td>SchoolFusion</td>
<td>NC Wise Owl</td>
<td>FirstClass</td>
<td>Castle Learning</td>
</tr>
<tr>
<td>SharpSchool</td>
<td>SuccessMaker</td>
<td>Fizz</td>
<td>Destiny/Follett</td>
</tr>
<tr>
<td>Angel</td>
<td>Waterford</td>
<td>iChat</td>
<td>Fast ForWord</td>
</tr>
<tr>
<td>eChalk</td>
<td>Learn 360</td>
<td>JupiterGrades/SnapGrades</td>
<td>PLATO</td>
</tr>
<tr>
<td>PB Works</td>
<td>NovaNet</td>
<td>LifeSize</td>
<td>Scholastic (incl SRI, SMI)</td>
</tr>
<tr>
<td>Dropbox</td>
<td>mClass</td>
<td>MyVRSpot</td>
<td>Teacher Tube</td>
</tr>
<tr>
<td>Ed Options</td>
<td>Renaissance Learning</td>
<td>Office Communicator</td>
<td>YouTube</td>
</tr>
<tr>
<td>Gaggle</td>
<td>Alexandria</td>
<td>VoiceThread</td>
<td>Apex Learning</td>
</tr>
<tr>
<td>Haiku</td>
<td>OdysseyWare</td>
<td>Wordpress</td>
<td>iTunesU</td>
</tr>
<tr>
<td>Learning Village</td>
<td>Classworks</td>
<td>ExamView</td>
<td>MediaCAST</td>
</tr>
<tr>
<td>LearnNC</td>
<td>Apex Learning</td>
<td>Final cut</td>
<td>NC Live</td>
</tr>
<tr>
<td>PD 360</td>
<td>Fast ForWord</td>
<td>HeadSprout</td>
<td>Orchard</td>
</tr>
<tr>
<td>Schoology</td>
<td>iStation</td>
<td>Inspiration/Kdspiration</td>
<td>Thinkfinity</td>
</tr>
<tr>
<td>Smart Suite</td>
<td>Learning.com</td>
<td>Intel Teach</td>
<td>ThinkGate</td>
</tr>
<tr>
<td>Adobe Acrobat Connect</td>
<td>Blue Dond</td>
<td>Khan Academy</td>
<td>AIMSweb</td>
</tr>
<tr>
<td>Aleks</td>
<td>Castle Learning</td>
<td>Learning.com</td>
<td>Animoto</td>
</tr>
<tr>
<td>Atomic Learning</td>
<td>Elements</td>
<td>Livebinders</td>
<td>Big Medium</td>
</tr>
<tr>
<td>Aventa</td>
<td>ExamView</td>
<td>Mathmatica</td>
<td>Blue Diamond</td>
</tr>
<tr>
<td>Blogger/Wordpress</td>
<td>Scantron</td>
<td>Media Master</td>
<td>BookFlix</td>
</tr>
<tr>
<td>BrainHoney</td>
<td>Scholastic (incl SRI/SMI)</td>
<td>Engrade</td>
<td>Britannica Online</td>
</tr>
<tr>
<td>CANSAS</td>
<td>TetraData (Follett)</td>
<td>My Access</td>
<td>C-MAPP</td>
</tr>
<tr>
<td>Destination Success</td>
<td>Academy of Reading / Math</td>
<td>My Big Campus</td>
<td>Capstone Library</td>
</tr>
<tr>
<td>Edublogs</td>
<td>Case21</td>
<td>PBS</td>
<td>Carnegie Learning</td>
</tr>
<tr>
<td>Elluminate</td>
<td>Classroom Manager</td>
<td>Rosetta Stone</td>
<td>CCC Video on Demand</td>
</tr>
<tr>
<td>First Class</td>
<td>Compass Learning</td>
<td>Orchard</td>
<td>CM595</td>
</tr>
<tr>
<td>Glogster</td>
<td>DiBELS</td>
<td>Carnegie Learning</td>
<td>CommonSenseMedia.org</td>
</tr>
<tr>
<td>IQity</td>
<td>NetTrekker</td>
<td>CMS95</td>
<td>CompassLearning</td>
</tr>
<tr>
<td>KidBlog</td>
<td>Office Communicator</td>
<td>Curriculum Associates</td>
<td>Curriculum Learning Associates</td>
</tr>
<tr>
<td>NetTrekker</td>
<td>Adobe Connect</td>
<td>EasyTech</td>
<td>EASI System</td>
</tr>
<tr>
<td>OnCourse</td>
<td>Atomic Learning</td>
<td>Education City</td>
<td>EasyTech</td>
</tr>
<tr>
<td>Open Office IT Academy</td>
<td>iChat</td>
<td>Google Apps</td>
<td>Ebsco Points of View</td>
</tr>
<tr>
<td>OpenGoo</td>
<td>InfoCenter</td>
<td>Harcourt Math</td>
<td>EduOptions</td>
</tr>
<tr>
<td>Plato</td>
<td>ooVoo</td>
<td>Imagine Learning</td>
<td>EducationCity</td>
</tr>
<tr>
<td>Quest Academy (eCollege)</td>
<td>RiverDeep</td>
<td>IXL Math</td>
<td>EduVision</td>
</tr>
<tr>
<td>Renaissance Place</td>
<td>SkillsBank</td>
<td>Jato</td>
<td>Elements</td>
</tr>
<tr>
<td>School In Sites</td>
<td>SkillsTutor</td>
<td>Kamico</td>
<td>Enchanted Learning</td>
</tr>
<tr>
<td>Smarthinking</td>
<td>TargetTeach</td>
<td>NCIH</td>
<td>eSchools</td>
</tr>
<tr>
<td>Study Island</td>
<td>TeachScape</td>
<td>Quia</td>
<td>Math Aligned</td>
</tr>
<tr>
<td>Teacher Studio</td>
<td>Testmate Clarity</td>
<td>Read180</td>
<td>metameetrics</td>
</tr>
<tr>
<td>Tetradata (Follett)</td>
<td>Turning Points</td>
<td>Reading Counts</td>
<td>My Access</td>
</tr>
<tr>
<td>EdOptions</td>
<td>Promethean Planet</td>
<td>Passport (Voyager U)</td>
<td>Edusoft</td>
</tr>
</tbody>
</table>
PART IV BUILD AND IMPLEMENTATION
Developing a Data Integration Service

Build and Implementation Model

Subphase 1: Data Integration

Subphase 2: Operational Data Store
Technical Implementation Plan Development

The Technical Implementation Plan will describe the technical details of how the DI system will be implemented in the context of a specific set of vendor products and platforms. The implementation must follow industry and product-specific document best practice guidelines. The DI Managed Service vendor will lead this effort with oversight from the DI Service Manager. The implementation plan will require approval from both the DI Service Manager and DI Policy and Governance before actual implementation is started. Several details of the technical implementation plan will inform the Service Plan that is being developed by the DI Service Manager.

Requirements of Technical Implementation Plan

In addition to the details/steps of how the DI system will be physically and logically implemented, the requirements below must be included in the implementation plan.

Infrastructure Specifications. The implementation plan must provide the necessary hardware and other infrastructure requirements, specifications, and recommendations needed for the particular product components of the DI system.

Authoritative Source System Analysis. The implementation plan must provide an analysis of the source systems that will be provided as input to the DI system. Any findings that impact the system should be described along with any recommendations from the DI Service Manager.

Approach. The implementation must provide details about strategies and general approach of the system implementation.

Data Profiling. A large portion of work in the implementation of any data integration system will be the data profiling exercise. The implementation plan must include the expanded data profile to be used in the DI system.

Software Maintenance. The implementation plan must describe the necessary software maintenance operations that are required to maintain a secure, high performing, highly available and sustainable system.

Infrastructure Maintenance. The implementation plan must describe the necessary hardware/infrastructure maintenance operations that are required to maintain a secure, high performing, highly available and sustainable system.

Data Maintenance. The implementation plan must describe the necessary data maintenance operations that are required to maintain a secure, high performing, highly available and sustainable system.
Best Practices. The DI system implementation must adhere to documented industry and product-specific best practices. The DI Managed Service should reference and highlight these best practices in the implementation plan as related to the actual DI system implementation.

Test Plans. The implementation plan must describe the system test plan and performance testing results. The results should correlate to system requirements and product performance profiles.

Associated Risks. The implementation plan must highlight any perceived risks associated with the DI System/Service.

Biographical Sketch. The implementation plan must provide a biographical sketch and qualifications of the current staff that would be involved in providing the platform components.

Commitments. The implementation plan must provide a description of other commitments by the DI Managed Service Provider as a metric of availability.

Personnel Requirements. The implementation plan must describe the internal personnel requirements to effectively manage the DI Platform initially and long-term. The personnel requirements should include each person’s role and responsibilities within the DI Platform along with necessary skills set needed.

Milestones and Deliverables. The implementation plan must describe the different milestones and deliverables for each phase of the implementation, deployment, and management of the DI Service.

Regulatory Compliance. The implementation plan must describe the audit standards level for any infrastructure hosting used for the DI System where appropriate. All DI System components are desired to be hosted in data centers that meet new SSAE 16 standards, including SOC1, SOC2, and SOC3\(^2\)\(^1\).

HA and DR. The implementation plan must describe the design and implementation aspects of how the DI system will perform and support High Availability and Disaster Recovery efforts. Also include Load Balancing, Fault Tolerance, and Scalability measures.

Change Management. The service plan must describe the DI Managed Service change management policies for all services rendered to DI Service Manager.
PART V

FULL PRODUCTION AND OPERATION
Production and Operations Model
Ongoing Support Model for LEAS

Ongoing Support Model for DPI

Ongoing Funding Model

Ongoing Integration for Data Sources

Ongoing Integration for Target Services

Ongoing Policy and Governance Model

Help Desk (Support) Escalation Process

Support – Cloud
(NC Education Cloud Help Desk)
Triaged to appropriate Cloud Service

Support – District
(Technology Director)

Support – School
(Teacher/Technology Facilitator)

Support – Service Manager

Support – Vendor
(Managed Service Provider)

K-12 User
Encounters Problem

*** It’s unclear at this time where a response would go, but likely to the school or district level (not the user)**
APPENDICES
Appendix A – Potential DI Future Work

This section describes a few additions, enhancements, etc. to the DI Service that are under consideration for future work efforts.

Data Federation

Data Federation provides a single virtual view of one or more source data files. When a business application issues a query against this virtual view, a data federation engine retrieves data from the appropriate source data stores, integrates it to match the virtual view and query definition, and sends the results to the requesting business application. By definition, data federation always pulls data from source systems by an on-demand basis. Any required data transformation is done as the data is retrieved from the source data files. The DI Service should support this functionality but will not take advantage of it during the initial phase of the project.

SOA and Message Queue Integration (Enterprise Service Bus)

The ESB and Message Queue functions add value to the DI Service/System. However, the State agencies supporting ESB components are not in a position to take advantage of this functionality currently. In the future, ESB could be added to enhance efficiency.

Complete Master Data Management

The DI Service will only utilize certain MDM functions such as data profiling, data matching, certain reports, etc. during the initial phase of the project. In the future, a wide range of other MDM functions could be added as the State agency builds resources.

Full Data Quality

The DI Service solution should support a wide variety of data quality functions. During the initial phase of this project, these functions will only be used for reporting and not actual cleaning of the data. Data quality issues should be reported and corrected at the source systems, following current data quality guidelines.
Operational BI and Operational Analysis

The DI Service should have the capability to analyze data for its own optimizations and functions. Using business intelligence and operations analysis functions are beyond the scope of this project and will be handled by external systems/tools/applications.

Local LEA Data Integration

The DI Service shall scale to support local LEA data movements and processing between LEA systems and between LEA and external systems. Due to the large breath of scope, the DI-MS will be limited to the source and target systems outline in earlier sections of this plan initially.
Appendix B – Regulatory and Compliance

The DI project will need to comply with and observe numerous federal/state regulatory and compliance rules due to its interaction with sensitive, personally identifiable information (PII) and various legislation regarding individuals with disabilities. Each of the following will potentially impact the DI Service at different levels. If applicable, the DI service must adhere to the guidelines establish by the following:

American Disabilities Act (ADA)\textsuperscript{22}

The American Disabilities Act (ADA) is a wide-ranging civil rights law that prohibits, under certain circumstances, discrimination based on disability. Disability is defined by the ADA as "a physical or mental impairment that substantially limits a major life activity." It was intended so that Americans with disabilities would be kept in the mainstream in terms of scientific and medical research and developments.

Children’s Internet Protection Act (CIPA)\textsuperscript{23}

The Children’s Internet Protection Act (CIPA) is a federal law enacted by Congress to address concerns about access to offensive content over the Internet on school and library computers. CIPA imposes certain types of requirements on any school or library that receives funding for Internet access or internal connections from the E-rate program – a program that makes certain communications technology more affordable for eligible schools and libraries. In early 2001, the FCC issued rules implementing CIPA.

Children's Online Privacy Protection Act of 1998 (COPPA)\textsuperscript{24}

The Children’s Online Privacy Protection Act, effective April 21, 2000, applies to the online collection of personal information from children under 13. The new rules spell out what a Web site operator must include in a privacy policy, when and how to seek verifiable consent from a parent and what responsibilities an operator has to protect children’s privacy and safety online.

Family Educational Rights and Privacy Act of 1974 (FERPA or the Buckley Amendment)\textsuperscript{25}

The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99) is a Federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

\textsuperscript{22} http://www.ada.gov/
\textsuperscript{23} http://www.fcc.gov/guides/childrens-internet-protection-act
\textsuperscript{24} http://www.coppa.org/comply.htm
\textsuperscript{25} http://www2.ed.gov/policy/gen/guid/fpc/ferpa/index.html
**Health Insurance Portability and Accountability Act (HIPAA) of 1996**[^26]

The Privacy Rule provides federal protections for personal health information held by covered entities and gives patients an array of rights with respect to that information. The Privacy Rule is balanced so that it permits the disclosure of personal health information needed for patient care and other important purposes.

**Individuals with Disabilities Education Act (IDEA)**[^27]

The Individuals with Disabilities Education Act (IDEA) is a United States federal law that governs how states and public agencies provide early intervention, special education, and related services to children with disabilities. It addresses the educational needs of children with disabilities from birth to age 18 or 21.

**Federal Information Security Management Act (FISMA)**[^28]

The E-Government Act (Public Law 107-347) passed by the 107th Congress and signed into law by the President in December 2002 recognized the importance of information security to the economic and national security interests of the United States. Title III of the E-Government Act, entitled the Federal requires each federal agency to develop, document, and implement an agency-wide program to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source.

**North Carolina Identity Theft Protection Act**[^29]

The North Carolina Identity Theft Protection Act of 2005 is a series of broad laws that was passed by the General Assembly of the U.S. state of North Carolina to prevent or discourage identity theft as well as guarding and protecting individual privacy.

[^26]: http://www.hhs.gov/policies/index.html
[^27]: http://idea.ed.gov/download/statute.html
[^28]: http://csrc.nist.gov/groups/SMA/fisma/overview.html
Appendix C – DI Research

This appendix will become a cumulative list of research done by the DI team to better understand the work of Data Integration. Also, including RFPs released by other organizations related to DI systems.

**DI Working Group.** The DI Working Group, which includes representatives from DPI, has been of tremendous value to this project. It has provided vast amounts of feedback and clarifications to the DI team. It has also given us a vast insight into the DPI environment that has continuously improved the DI plan.

**Vendor Information Sessions.** Many vendors spent hours and days with us. The vendors provided an invaluable amount of information around the Data Integration spectrum and greatly informed our planning process.

**LEA Site Visits.** Several LEA districts graciously hosted the DI Team for visits to discuss their data integration systems. These visits were in addition to the larger statewide site interviews done by the NC Education Cloud project. These in-depth sessions have been and will continue to be immensely valuable to progress. The opportunity to meet the people and understand the systems in action greatly enhanced our research.

**Various RFPs.** Much useful information has been garnered from previously released RFPs from a variety of sources, including the Delaware Department of Education, Denver Public Schools, North Carolina ITS, New York City, Swarthmore, Tennessee, Wisconsin, MnSCU, North Carolina Instructional Improvement System, North Carolina Identity and Access Management, etc. We believe these will enhance any procurements associated with our DI project.
Appendix D – Requirements for Integrating Cloud Applications

This appendix will serve as a reference and requirements guide for vendors, cloud–based service providers, and other RttT project teams, for planning purposes around data integration management.

The following are applicable to all models:

Policy and Governance. Each cloud application, resource, or service that integrates with the central DI system will require a certain level of policy and governance that defines how and what data is exchanged. Personally identifiable information (PII) will be provided for some applications, and service level agreements (SLAs) or contracts need to be in place to enforce confidentiality and data disclosure policies.

Detailed Technical Specifications. When the time comes to integrate cloud-based applications, resources, or services with the DI system, a more detailed set of technical specifications will be needed and provided. Some of these specifications cannot be documented until the central DI system actually exists, but sufficient planning can be done based on the information provided below.

Security. The DI system will require the secure transfer of the user identity data to and from itself. Whether from the source systems to the DI system, or between the DI system and cloud applications, all user data will be encrypted during transfer. Data transferred internal to the DI System will also be over encrypted channels where necessary. If user identity data is provisioned into a cloud application, the policy and governance section will define how data is managed throughout the process, along with any SLA or contractual obligations related to its storage and protection at the vendor site.

“Currency” of User Data. Updates made to user data in the Source Systems will be reflected in the DI System as soon as it is available and can be processed. This will vary greatly depending on the availability and format of the feeds. Initially, it is expected that updates will take place nightly, with full database extracts from the source systems being processed in a batch mode. This will naturally result in up to a 24-hour delay in user updates being made available to the DI System. A more “real-time” availability of data changes would be possible if the source systems would allow direct connections to a mirror set of the data or could provide deltas of changed data only, on a more frequent basis. This may be possible once the DI system is in production, agreements and SLAs are in place, and more is known about the entire process. As previously stated, it is important to get the data to the target services at the “right-time.”
Consuming/Acquiring Source Data

**Model 1: Data Integration Service pulls from Source Systems (Preferred Model)**
The DI service will support a wide array of heterogeneous and disparate source systems. This model describes how the data integration service will connect to the source systems and retrieve (pull) necessary data. The data integration service must have an agreement with the source systems as to how it intends to access the data, what data will be accessed/acquired, and at what speed/frequency. This agreement will be driven in part by the target services that are requesting data sets.

**Model 2: Source Systems push to Data Integration Service**
The DI service will provide a landing area (or staging area) for source systems to push data to in the event that Model 1 is not supported. A standard format will be needed for this model. The landing area will support an encrypted file system for security purposes. An example protocol for model 2 could be SFTP.

---

**Data Integration - Consuming/Acquiring Source Data**

Model 1: Data Integration Service pulls from Source Systems (e.g. API call, DB query, SFTP)

Model 2: Source Systems push to Data Integration Service (e.g. SFTP)
Target Data Delivery

**Model 3: Data Integration Service pushes to Target Systems**
The DI service will support a wide array of heterogeneous and disparate source systems. This model describes how the data integration service will connect to the target systems and deliver (push) necessary data. The data integration service must have an agreement with the target systems as to how it intends to consume the data, what data will be accessed/acquired, and at what speed/frequency. This agreement will be driven in part by the source systems that are granting access to the requested data sets.

**Model 4: Target Systems pull from Data Integration Service via ODS (Preferred Model)**
The DI Service will supply a central data repository, the Operational Data Store, which source data is replicated to. This model describes how target services can query the ODS for the needed data. This can be in real-time or via batch extract depending on the needs of the target service. As with the other models, the data integration service must have an agreement with the target systems as to how it intends to consume the data, what data will be accessed/acquired, and at what speed/frequency.
Appendix E – DI Project Assumptions, Risks, and Dependencies

This section will describe Assumptions, Risks, and Dependencies of the DI project. Risks that are described could become significant obstacles and pitfalls for the project. These will be captured and reviewed during a Monthly Risk Assessment (of the plan).

Assumptions
One key assumption, upon which all of this DI work is based, is that over time LEAs will be moving to cloud applications and reducing their use of locally based services.

Risk Analysis Table

<table>
<thead>
<tr>
<th>Category: Risks are grouped under Category Heading (e.g. RFP Process, Project - Development, Project - Production, People Issues, Security &amp; Compliance).</th>
</tr>
</thead>
</table>

Rating: Color indicator designating the amount of risk based on the formula* (below).

Risk Description: A brief description of the risk or dependency in enough detail to understand the issue and provide appropriate preventative measures or mitigation.

Likelihood: How likely it is that the risk will evolve into a real problem. The likelihood may change over time; something that is unlikely to be a problem at the start of a project may become very likely when the due date is approaching and the risk has not yet been avoided. A simple ranking of high, medium, low is used.

Impact: What is the impact if the risk becomes a real problem. Impact is a bit subjective, but should address the overall project. For example, a high impact risk is one that could cause the entire project to be canceled or significantly delayed. A risk of low impact might mean increased cost or a small impact to project schedule. Again, in this case a simple rating system of high, medium and low is used.

Prevention Plan: This is what will be done to ensure that the risk does not become a problem. For dependencies, this might include communicating with the supplier early and often, tracking interim milestones, etc. For technical risks it might mean doing early prototype work, or adding subject-matter experts to the team. This is “risk avoidance”.

Contingency Plan: This is what will be done in the event the risk escalates into a problem – in other words, the prevention plan was not successful.

Narrative: This contains a more detailed description of the risk, along with possible examples.

Likelihood & Impact Scale: Low=1, Medium=2, High=3

*Adopted from http://blogs.oracle.com/tacticalleadership/entry/know_what_you_re_doing3
Developing a Data Integration Service

*Formula for Red, Yellow, Green Risk Color:
Likelihood x Impact = Risk factor. 1-3 (Green), 4-5 (Yellow), 6-9 (Red)*

<table>
<thead>
<tr>
<th>Rating</th>
<th>Risk Description</th>
<th>Likelihood</th>
<th>Impact</th>
<th>Prevention Plan</th>
<th>Contingency Plan</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFP BAFO Too High</td>
<td>DI Service needs to be broken into very specific phases/stages with what</td>
<td>High</td>
<td>Med</td>
<td>If all prevention plan items have been followed, contingency would be to either</td>
<td>Responses to RFP for DI Service Solution come back with too high a cost (possibly due to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>functionality will be delivered in each phase. Early phases must stand on their</td>
<td></td>
<td></td>
<td>further reduce requirements, or look at some lower tier vendors</td>
<td>complexity of product/solution, extent of functionality/features, scope creep beyond</td>
<td></td>
</tr>
<tr>
<td></td>
<td>own (deliver basic functionality)</td>
<td></td>
<td></td>
<td></td>
<td>specified requirements)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Later stages can be “optional” functionality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Respondents need to price each phase/stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineffective</td>
<td>Need to really do our due diligence in the vendor evaluation phase. Try and</td>
<td>Med</td>
<td>High</td>
<td>Demand the primary respondent replace the SI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Product</td>
<td>get to know the actual people who will be doing the work, before choosing the</td>
<td></td>
<td></td>
<td></td>
<td>VERSUS What’s NEEDED:</td>
<td></td>
</tr>
<tr>
<td>Architect (Systems</td>
<td>vendor</td>
<td></td>
<td></td>
<td></td>
<td>- Incompetent (skillset, knowledge)</td>
<td></td>
</tr>
<tr>
<td>Integrator)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Ineffective Project Management</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Personality issues</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Lack of collaboration (no knowledge transfer)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VERSUS What’s NEEDED:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Understanding of Products and Functionality</td>
<td></td>
</tr>
<tr>
<td>Target Systems</td>
<td>Confirm (before RFP?) with rest of Cloud Team on which target(s) should be</td>
<td>Med</td>
<td>High</td>
<td>Make a best effort to list our expected top DI Service recipients. If ODS model is</td>
<td>Other Cloud Services needing DI Services data aren’t identified to enable “target</td>
<td></td>
</tr>
<tr>
<td>(Services) not identified</td>
<td>first priority. Then speak directly to these targets in the RFP.</td>
<td></td>
<td></td>
<td>followed, unknown target services can be configured to pull appropriate data more</td>
<td>systems” to be listed in the RFP – may impact solution or at least “cost”</td>
<td></td>
</tr>
<tr>
<td>prior to RFP release</td>
<td></td>
<td></td>
<td></td>
<td>easily. Attempt to “lock in” a cost for connecting similar Target Services.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Target Systems (Services) not identified prior to RFP release

Other Cloud Services needing DI Services data aren’t identified to enable “target systems” to be listed in the RFP – may impact solution or at least “cost”
### DI Service Configuration & Implementation

<table>
<thead>
<tr>
<th>Risk Description</th>
<th>Likelihood</th>
<th>Prevention Plan</th>
<th>Contingency Plan</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unreliable Authoritative Data Sources</strong></td>
<td>High</td>
<td>Continued communication with all stakeholders. (Are there data steward(s) who are willing and able to work with us to facilitate proper access to and interpretation of any needed data?) Governance Board/Committee to review? Messaging to LEA community needs to include info on their data responsibilities. Should we begin formalizing the [change management] committee now?</td>
<td>Get data needed directly from LEAs? Technical committee / cloud team would be the fallback for enforcing Change Management</td>
<td>Is the input data quality absolutely horrible, irreparably so?</td>
</tr>
<tr>
<td>- Data Integrity</td>
<td>High</td>
<td></td>
<td></td>
<td>What if input data sources stop being available? Or if they’re corrupted?</td>
</tr>
<tr>
<td>- Data Access</td>
<td>High</td>
<td></td>
<td></td>
<td>When the input data sources change, will we be adequately notified?</td>
</tr>
<tr>
<td>* Frequency</td>
<td>High</td>
<td></td>
<td></td>
<td>What if the input data doesn’t have everything we need? (e.g. if don’t have course enrollment / teaching info.)</td>
</tr>
<tr>
<td>* Delivery</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Sets Availability</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Key Attributes</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Ineffective Change Management (affecting Data Access or Data Integrity)</td>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Unreliable Support for Data</strong></td>
<td>High</td>
<td>Need to approach Source System “owners” now to identify the authorized, capable Point of Contact who will help us through data issues. LEAs will need to update / clean data, also web page(s) with instructions</td>
<td>This one doesn’t really have a contingency. It requires the Governance committees to impose stricter requirements on the data stewards to provide appropriate support.</td>
<td>DPI, Wake, Charlotte, NCSSM, must provide input data consistently, along with advice and assistance on that data. “One off” systems such as NCSSM, standalone HR systems, etc. may also need to provide data.</td>
</tr>
<tr>
<td><strong>Scope Creep</strong></td>
<td>Medium</td>
<td>Stay focused on clear, measurable, achievable, and beneficial functionality (the 80/20 rule)</td>
<td>Reprioritize to focus on the absolute bare essential requirements only.</td>
<td>Scope Creep – attempting to provide too much functionality, too many “bells &amp; whistles”, political pressure to add more than what can be funded</td>
</tr>
<tr>
<td><strong>Use/requirement of Nonstandard Connector/Adapter</strong></td>
<td>Low</td>
<td>Make sure these issues are addressed with questions in the RFP or Q&amp;A during selection Ensure that a “connector” is needed for a resource</td>
<td>One option might be to develop (program) the connector “in-house” rather than use the vendor.</td>
<td>- Lack of appropriate connectors</td>
</tr>
<tr>
<td><strong>Missed Key Requirement(s) from Stakeholders (LEAs and DPI)</strong></td>
<td>Low</td>
<td>Continued communications with LEAs and DPI LEA,DPI internal communication (top to bottom, regional representative to LEAs)</td>
<td>Require funding to accommodate changes, additional features</td>
<td>Requirement may impact the way the DI Service works, or the priority of functionalities implemented.</td>
</tr>
</tbody>
</table>

**Notes:**
- High: High Likelihood, High Impact
- Medium: Medium Likelihood, Medium Impact
- Low: Low Likelihood, Low Impact

**Service Configuration & Implementation:**
- Developing a Data Integration Service
- Service Configuration & Implementation
- Risk Management
- Prevention Plans
- Contingency Plans
- Narrative

**Additional Notes:**
- Ensure that a “connector” is needed for a resource
- One option might be to develop (program) the connector “in-house” rather than use the vendor.
### Communication & People Issues

<table>
<thead>
<tr>
<th>Rating</th>
<th>Risk Description</th>
<th>Likelihood</th>
<th>Impact</th>
<th>Prevention Plan</th>
<th>Contingency Plan</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ineffective/Inadequate Communication</td>
<td>High</td>
<td>High</td>
<td>Regularly scheduled status meetings between these groups should specifically address this point, as well as written status updates.</td>
<td>This is a leadership issue that the Governance Committees must address if it becomes a problem.</td>
<td>Among and between stakeholders, technical, management, data stewards, and DI Service “roles”, etc.</td>
</tr>
<tr>
<td></td>
<td>Ineffective Governance Boards/Committees</td>
<td>High</td>
<td>High</td>
<td>This gets back to the need for project “champions” throughout the governance structure, to encourage timely responses to Governance and Policy queries.</td>
<td>Where do you go if the top-level leaders are not effective?</td>
<td>Failure of Governance committees to determine/make policy</td>
</tr>
<tr>
<td></td>
<td>Poor Transitions / Handoffs</td>
<td>Medium</td>
<td>Medium</td>
<td>Set expectations early on and build in clear, regular, measurable checkpoints! Need to really do our due diligence on the primary vendors -- and hold them accountable in case of problems. If Vendor is remote, perhaps the regular checkpoints should be F2F?</td>
<td>Might have no other choice than to reduce the deliverables of early phases.</td>
<td>Initial DI Managed Service development phase could take too long</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vendor sub-contractors could be in over their heads</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Possible vendor bait and switch (they might bring in a SI team that’s different?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Communication with DI Managed Service (non-responsiveness) -&gt; if SI is remote, how will that impact us?</td>
</tr>
</tbody>
</table>
### Developing a Data Integration Service

<table>
<thead>
<tr>
<th>Rating</th>
<th>Risk Description</th>
<th>Likelihood</th>
<th>Impact</th>
<th>Prevention Plan</th>
<th>Contingency Plan</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sustainability</td>
<td>Medium</td>
<td>Medium</td>
<td>Need to plan starting now for the end of RttT funding - what is the model after that? Functionality provided needs to match funding – if funding is reduced, then functionality might need to be restricted Cost of additional functionality, apps, connectors/adapters - will need to be funded by either DPI, LEAs (for small group of beneficiaries), State, App Vendor, etc.</td>
<td>Each of these concerns needs to be addressed. The contingency is to cut back on services or find additional funding. Some of these might need to be dealt with as they arise, and do whatever it takes to maintain the DI Service – this assumes that Cloud project and Shared Infrastructure and Services are successful.</td>
<td>We MUST develop a sustainable plan to address the critical cost issues!! Lack of funding Inability to complete/expand DI Service Later phases Loss of funding Nationals/State economic downturn (essentially same as above) Vendor Product Ongoing Cost/Increased Cost Vendor buyout Product change Lack of mitigation strategies</td>
</tr>
</tbody>
</table>
## Security and Compliance

<table>
<thead>
<tr>
<th>Rating</th>
<th>Risk Description</th>
<th>Likelihood</th>
<th>Impact</th>
<th>Prevention Plan</th>
<th>Contingency Plan</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Security of PII Data (Target Services, Apps)</td>
<td>Low</td>
<td>High</td>
<td>Contracts and SLA with Vendors need to stipulate requirements and responsibilities of any data they store – as well as liability in the event of an incident</td>
<td>If contracts are written carefully, then litigation is the contingency to non-compliance.</td>
<td>Service providers -- can we trust them to protect PII? Will they responsibly consume assertions?</td>
</tr>
<tr>
<td></td>
<td>Breach of Systems &amp; Data, Password Compromise (DI System - Internal)</td>
<td>Low</td>
<td>High</td>
<td>Breaches and sharing of data can and do happen. Having a well thought out (and tested) plan for remediation of any security issue is critical to keeping data secure.</td>
<td>Having a detailed plan for addressing any type of breach or data release would be the appropriate contingency.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Intentional or Unintentional Exposure or Release of PII Data</td>
<td>Low</td>
<td>Medium</td>
<td>In either case, this involves a well thought out and detailed mitigation plan. If the release was intentional, procedures for disciplinary action by the appropriate authority should be in place and communicated through normal channels.</td>
<td>SIEM Tool</td>
<td>Compliance risks – what if we accidentally (or intentionally) violate one of the state or federal privacy regulations (FERPA, HIPAA)? If so, how to fix? What would be the ramifications?</td>
</tr>
</tbody>
</table>

### Dependencies

Key dependencies, such as timely access to accurate data sources, have been outlined in the above risks table. If critical dependencies are not realized, this project’s success will be impeded.
Appendix F - Frequently Asked Questions

FAQs are gathered from LEA Site Interviews, LEA conference presentations, webinars, and other communication mediums between stakeholders and NC Education Cloud team. They will be listed below and have the best response available with the current information and plans.

**Q. What reliability can the LEAs expect for the DI system?**

A. One of the topics during vendor discussions was that of Service Level Agreements (SLAs) and how many "nines" of availability the DI system should have. We plan for the DI system to operate continuously, with any expected downtime scheduled well in advance and during off-hours, and communicate to all stakeholders. Through discussions with the DI Working Group members, it was determined that 99.9% uptime would be sufficient. This would allow no more than 2.19 hours per quarter of unexpected downtime.

**Q. Who will oversee the DI system when it is finished?**

A. The cloud team is working to build a sustainable support model for all cloud based services. The current recommendations are to form an education service agency and support network.

**Q. Who and how many people will run the state help desk for DI?**

A. The Cloud team is currently developing a help desk model that encompasses support for the DI components as well as other services driven by the cloud project. The team is in the early stages of requirements gathering to help define LEA support needs surrounding the new cloud services model.

**Q. Is the DI service a replacement for MFTS?**

A. The DI Service is not intended to be a replacement for the MFTS service. The DI solution has a much bigger scope but has the capability to complement MFTS as needed.
Appendix G – Proposed Policies, Regulations and Rules (PRRs)

These questions will eventually become more formalized PRRs with structure and process in the larger NC Education Cloud context. We will be proposing PRRs in response to these simple questions that need to be answered in order to make provisioning and workflow decisions.

Business Operations PRRs

1. Who is involved in contracts (outside of DPI written or purchased applications)?

2. Can LEAs enter into their own agreements with vendors, and if so – can they also authorize the release of student/staff data?

   A: LEAs can certainly enter into their own agreements with vendors, however it would not be guaranteed the DI system would release any data to those external vendor applications. The DI governance body, in collaboration with the DI technical team, would need to approve and then implement releasing data from the DI system. If this approval and implementation does not occur, the LEA would need to interact directly with the vendor to provide any necessary data to that service.

3. What happens to the DI system (service) after the NC Education Cloud program is over? (How is the DI project sustained to continue providing value to stakeholders?)

4. Who oversees the DI project after the Cloud Project is completed?

5. Will LEAs be allowed to interact directly with the DI Service? If so, how would they request information, access, and resources?

6. How does a vendor / service request access to the data? Workflow? Process?

7. Who maintains list of administrators that get alerts/notifications from DI-MS?

Data Governance PRRs

Governance is critical for any data integration solution. In order for data to be transformed into information - it must have meaning. For example:

1. What do specific artifacts (e.g., data elements) mean if you’re an educator?

2. Who owns it?
3. How do I use it?

4. Are there any rules that I should know about?

5. What are valid values?

6. When was it last extracted (e.g., how fresh is the information)?

7. Where did the information come from (e.g., Data was extracted from Mecklenburg Lawson HR system)?

8. What are the business requirements?

9. What datasets need real-time delivery?

10. What standards are there for preferred interfaces for data integration scenarios?

11. What schemas are to be used for exchanging data?

12. What is the process for getting access to the authoritative data sources for students, employees/staff and service providers?

13. What Regulatory and Compliance Rules impact or apply to the DI System?

14. Who owns the user data? (LEA, DPI, Parent, Staff member)

   A: LEAs “own” the user data on students and staff members. They also approve, or “sponsor” the data on guest members such as parents, volunteers, etc., so in a sense the LEAs own that data as well. The DI system does not own any of the data; it simply has custody of it on behalf of the LEAs and DPI.

15. What data (elements, attributes) will be made available to the ODS?

16. Will parent data be self-asserted (captured via some other mechanism), or only provided via NCWISE?

   A: Parent data in NCWISE will likely not be used until it becomes a more reliable source. Instead a self-registration (guest) system combined with LEA-approval will probably be used.

17. What data is authoritative for Parents that are also Staff (the Guest or HR System)?

18. Once the data is transferred to the “Cloud”, does the ownership change?
A: No, however, the custodian may change, e.g. from the DI system to a vendor application provider. External service providers will be mandated to adhere to strict safeguards to protect the privacy of any such transferred data.

**Data Source PRRs**

1. Who are the data steward contacts for each source? Business Steward? Technical Steward? System Administrator?
2. What documentation is available about source systems?
3. Who authorizes access to the source systems?
4. What source systems are authoritative for what data?
5. What certification process do the source systems go through to verify they are compatible with DI consumption models?

**Data Transformation PRRs**

1. What are the various transformations needed between the source and target systems?

**Data Quality PRRs**

1. What fields are critical?
2. What is the exception threshold?

**Data Targets PRRs**

1. What data does each target service need?
2. What format does the target service need the data in?
3. What is the consumption process of the target service?
4. What is the approval process to authorize data release?
5. What certification process do the target services go through to verify they are compatible with DI data delivery models?

Operational Data Store PRRs

1. What is the data model?

2. What API functions are available?

3. What type and age of data will the ODS hold?
Appendix H – MFTS File Naming Convention

Each file submitted by an authoritative publishing system will contain the following attributes as part of the file name. This data is critical to both the correct processing and identification of files and can contain no blanks or white space.

File ID Construction Convention

[Customer ID]-[File Creation DateTimeGroup]-[Customer File ID]-[Movement Group ID]-[File Type ID]-[Reference ID]-[Customer String].[Extension]

1. Customer ID – 4 position Alphanumeric Identifier issued by the system to registered customers. The customer ID represents a logical endpoint node in the MFTS network. Multiple logical endpoints may exist on a single physical server – for example in the case of an application-hosting server with multiple applications residing on it. Each application that participates in the MFTS network as an endpoint must have its own Customer ID. Files distributed by the MFTS are delivered on the basis of one copy per endpoint.

1. File Creation DateTimeGroup (DTG) – CCYYMMDDhhmms – century year month day hour minute seconds

1. Customer File ID – 2 position Numeric Identifier determined by the customer. Needs to be unique within a DTG. Range 0-99. This number is always represented as 2 digits, eg. 00, 01, 02, …99

1. Movement Group ID – 4 position alphanumeric identifier issued by the system when a transfer configuration record is created. This ID tells the system which transfer record is applicable to this file allowing it to route the file correctly. This effectively provides the addressing instructions for the file in the absence of an address label and makes multiple addressees possible. This element is not guaranteed to remain consistent throughout MFTS processing so consumer applications should not use it to control processing.

1. File Type ID – A 5 position alphanumeric identifier for use by the target system in processing the file. This value will typically identify the file as being a specific publication that should have meaning to consuming endpoints.

1. Reference ID - Optional ID that combines the Customer ID, File Creation DateTimeGroup, and Customer File ID of a previous file that this file references. If there is no previous file to reference, this group should contain nothing which will result in 2 segment separators being created next to each other.

1. Customer String – A customer created string to help humans looking at logs, directory entries, etc. group and locate files manually. Does not need to be unique at any level but must insure that the total length of the filename does not exceed 128 characters including the extension. The delimiter for this string must be included even if the string is null.

1. Extension – A 3 position Alpha Identifier extension that identifies the type of file. Acceptable values are: txt, csv, dat, xml. Files that are in the process of being written must use the ‘.tmp’ extension to identify them as in process. MFTS pickup routines will not collect any files ending in ‘.tmp’.
WARNING:
Your process must write files with the ‘.tmp’ extension during transfer to the MFTS, then rename the files to one of the acceptable data file extensions. Failure to follow this procedure could result in data corruption including partial record collection and record loss.

Definitions follow:

- Alphanumeric Identifier: (a-z), (A-Z), (0-9), ( _ ) (underscore)
- Numeric Identifier: (0-9)
- Segment separator: - (dash)
- Extension separator: . (period)
- Alpha Identifier: (a-z) – generally lower case

Examples:

Example 1:

[Customer ID] = UIDS
[File Creation DateTimeGroup] = 20080329144533
[Customer File ID] = 01
[Movement Group ID] = UID2
[File Type ID] = UIDR
[Reference ID] = NCWS2008032912220301
[Customer String] = UID Response
[Extension] = csv

Filename: UIDS-20080329144533-01-UID2-UIDR1-NCWS2008032912220301-UIDResponse.csv

Example 2:

[Customer ID] = NCWS
[File Creation DateTimeGroup] = 20080115122243
[Customer File ID] = 06
[Movement Group ID] = 1001
[File Type ID] = UIDSB
[Reference ID] = [NULL]
[Customer String] = Burns_High_School
[Extension] = txt

Filename: NCWS-20080115122243-06-1001-UIDSB--Burns_High_School.txt
Appendix I – NCWISE Student Demographics Extract Data Definitions

Below are sample extracts from NCWISE using MFTS. These give a sample of the type of data fields that will need to be profiled and transformed by the DI-MS.

Purpose:  
**Daily export of Student Demographics data from NCWISE**

Extract Filename:  
1NCW-20110310103414-00-2602-SUREQ--NCWISE.zip  
File Type: zip  
Delimiter: Tab

Last Updated:  
5/25/12 5:04 PM  
MGID: 2602  
Escape: **none**

Sample Filename:  

Population:

- Extract contains Active Student Records from NCWISE (WITHDRAW_DATE is null, and REGISTRATION_DATE is not null)

Fields:

<table>
<thead>
<tr>
<th>Col</th>
<th>Field</th>
<th>Size</th>
<th>Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PUPIL_NUMBER</td>
<td>9</td>
<td>Alpha</td>
<td>NCWISE Student ID. Pupil number is generated by the state and is unique to each student. Pupil Number contains leading zeros.</td>
<td>012345678</td>
</tr>
<tr>
<td>2</td>
<td>LAST_NAME</td>
<td>30</td>
<td>Alpha</td>
<td>(LEGAL_NAME) Student legal last name</td>
<td>Smith</td>
</tr>
<tr>
<td>3</td>
<td>FIRST_NAME</td>
<td>30</td>
<td>Alpha</td>
<td>Student First name</td>
<td>John</td>
</tr>
<tr>
<td>4</td>
<td>SECOND_NAME</td>
<td>40</td>
<td>Alpha</td>
<td>Middle/Maiden Name</td>
<td>Andrews</td>
</tr>
<tr>
<td>5</td>
<td>LAST_NAME_SUFFIX</td>
<td>4</td>
<td>Alpha</td>
<td>Student’s last name suffix.</td>
<td>Jr., II, III, etc</td>
</tr>
<tr>
<td>6</td>
<td>BIRTH_DATE</td>
<td>10</td>
<td>Alpha</td>
<td>Student birth date MM/DD/YYYY</td>
<td>03/15/2001</td>
</tr>
<tr>
<td>Field</td>
<td>Length</td>
<td>Type</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------</td>
<td>------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHYSICAL_STREET_ADDRESS</td>
<td>37</td>
<td>Alpha</td>
<td>Student physical street address. (STREET_NUMBER space STREET_NAME)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APARTMENT</td>
<td>10</td>
<td>Alpha</td>
<td>Student apartment number or letter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSTAL_BOX</td>
<td>6</td>
<td>Alpha</td>
<td>Student Post Office Box number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITY</td>
<td>30</td>
<td>Alpha</td>
<td>The city in which the student lives. (MUNICIPALITY_DESC from MUNICIPALITIES table).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STATE</td>
<td>4</td>
<td>Alpha</td>
<td>The standard two-letter state postal abbreviation for the state in which the student lives. Note: NCWISE field size is 4. (PROVINCE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZIP_CODE</td>
<td>10</td>
<td>Alpha</td>
<td>The postal zip code for the student address. (POSTAL_CODE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAILING_ADDRESS</td>
<td>150</td>
<td>Alpha</td>
<td>The address to which correspondence is mailed. This address can be the same as the property address, or it may be different.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHONE_NUMBER</td>
<td>10</td>
<td>Alpha</td>
<td>The home phone number of the student. (PHONE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRADE</td>
<td>2</td>
<td>Alpha</td>
<td>Grade of the student upon admission into the school</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOME_LANGUAGE_SURVEY</td>
<td>1</td>
<td>Alpha</td>
<td>Home Language Survey has been returned. Values: Y, N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANGUAGE_CODE</td>
<td>5</td>
<td>Alpha</td>
<td>Home Language Code, primary 116erive116e code. See LANGUAGE_CODES table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANGUAGE_DESC</td>
<td>30</td>
<td>Alpha</td>
<td>Home Language Description, primary language. From LANGUAGE_CODES table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTY</td>
<td>6</td>
<td>Alpha</td>
<td>Code for county of residence. See COUNTY_CODES table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COUNTY_DESC</td>
<td>50</td>
<td>Alpha</td>
<td>County description of residence. From COUNTY_CODES table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CITY_OF_BIRTH</td>
<td>30</td>
<td>Alpha</td>
<td>Student city of birth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STATE_PROV_OF_BIRTH</td>
<td>3</td>
<td>Alpha</td>
<td>The standard two-letter state postal abbreviation for the state in which the student was born. NOTE: NCWISE field size is 3.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** May contain Student’s Social Security Number, security precautions should be taken if accessing this field.
## Developing a Data Integration Service

<table>
<thead>
<tr>
<th>Field</th>
<th>Length</th>
<th>Type</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUNTRY_OF_BIRTH_CODE</td>
<td>3</td>
<td>Alpha</td>
<td>Country Code where student was born (COUNTRY_OF_BIRTH). See COUNTRIES table</td>
<td></td>
</tr>
<tr>
<td>COUNTRY_OF_BIRTH_DESC</td>
<td>30</td>
<td>Alpha</td>
<td>Country name where student was born. (COUNTRY_DESC from COUNTRIES table)</td>
<td>Botswana</td>
</tr>
<tr>
<td>MOTHERS_MAIDEN_NAME</td>
<td>40</td>
<td>Alpha</td>
<td>Not populated from NCWISE data, repository no longer needs this field.</td>
<td>NULL</td>
</tr>
<tr>
<td>NCWISE_LEA</td>
<td>9</td>
<td>Alpha</td>
<td>NCWISE Board/District/LEA identifier, up to 3 characters. (BSID from SCHOOLS table matched via SCHOOL in STUDENTS table)</td>
<td>57</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>9</td>
<td>Alpha</td>
<td>NCWISE numeric school code combines LEA code and school. From STUDENTS table</td>
<td>57057</td>
</tr>
<tr>
<td>LEA_SCHOOL_CODE</td>
<td>6</td>
<td>Alpha</td>
<td>DPI school designation including LEA and school number. School Code definition FFFLLL – EDDIE Format: FIRST 3-DIGITS FFF: 001-009 Regional Education – Service Alliances/Consortia 01A-995 LEAs (When all positions are numeric) Charter schools (When first 2 positions are numeric and 3rd position is alphabetic) Special LEAs: 209 – Cherokee Central 269 – Fort Bragg 679 – Camp Lejeune 997 – Dept of Human Resources 998 – Office of Juvenile Justice LAST 3-DIGITS LLL: 000 for LEAs and charter schools 300 or greater for schools</td>
<td>60G000</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

<table>
<thead>
<tr>
<th></th>
<th>Column Name</th>
<th>Length</th>
<th>Type</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>LEA_CODE</td>
<td>3</td>
<td>Alpha</td>
<td>LEA Code definition EDDIE Format:</td>
<td>60G</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>001-009 Regional Education – Service Alliances/Consortia</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>01A-995 LEA Code if all positions are numeric</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Charter school Code if first 2 positions are numeric and 3rd position is alphabetic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Special LEAs:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>209 – Cherokee Central</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>269 – Fort Bragg</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>679 – Camp Lejeune</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>997 – Dept of Human Resources</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>998 – Office of Juvenile Justice</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>SCHOOL_CODE</td>
<td>3</td>
<td>Alpha</td>
<td>3-DIGITS EDDIE Format:</td>
<td>000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>000 for LEAs and charter schools</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300 or greater for schools</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>GENDER</td>
<td>1</td>
<td>Alpha</td>
<td>(SEX)</td>
<td>M/F</td>
</tr>
<tr>
<td>33</td>
<td>MINORITY_CODE</td>
<td>3</td>
<td>Alpha</td>
<td>Currently, race/ethnicity in NC WISE is simply the old minority_code_1 field from the students’ table.</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>But, because of new Federal requirements, the race/ethnicity value has been updated to come from a different area of NC WISE and it is a more complex value.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>This field uses function get_race_fnc to derive the new Minority Code value, which may be different in some cases from the old MINORITY_CODE_1 value.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Values: AMIN=1, ASIA=2, HISP=3, BLCK=4, WHTE=5, MULT=6, PACI=7</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>MINORITY_DESC</td>
<td>30</td>
<td>Alpha</td>
<td>This field uses function get_race_fnc to derive the new MINORITY_DESC values.</td>
<td>White</td>
</tr>
<tr>
<td>35</td>
<td>HISPANIC_ETHNICITY</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENTS table column MINORITY_CODE_1, if ‘3’ return Y.</td>
<td>Y/N</td>
</tr>
<tr>
<td>36</td>
<td>AMIN_IND</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENT_HUMAN_RACES table column HUMAN_RACES_CODE column, if 1 return Y. (American Indian/Alaskan Native)</td>
<td>Y/N</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>ASIAN</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENT_HUMAN_RACES table column HUMAN_RACES_CODE column, if 2 return Y.</td>
</tr>
<tr>
<td>38</td>
<td>BLACK</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENT_HUMAN_RACES table column HUMAN_RACES_CODE column, if 4 return Y.</td>
</tr>
<tr>
<td>39</td>
<td>WHITE</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENT_HUMAN_RACES table column HUMAN_RACES_CODE column, if 5 return Y.</td>
</tr>
<tr>
<td>40</td>
<td>HAW_PI</td>
<td>1</td>
<td>Alpha</td>
<td>Based on NCWISE STUDENT_HUMAN_RACES table column HUMAN_RACES_CODE column, if 7 return Y. (Hawaiian/ Pacific Islander)</td>
</tr>
<tr>
<td>41</td>
<td>LEGAL_RES_DISTRICT_NUMBER</td>
<td>9</td>
<td>Alpha</td>
<td>Displays the LEA to which the student is zoned based on the student residence address. This functionality is only available if the LEA has selected this field for use in the Company File and the Board/Districts Table is populated.</td>
</tr>
<tr>
<td>42</td>
<td>PREVIOUS_SCHOOL_NUMBE R</td>
<td>9</td>
<td>Alpha</td>
<td>Student’s previous school/district information populates if a student moves from one NC WISE school to another.</td>
</tr>
<tr>
<td>43</td>
<td>DIPLOMA_TYPE</td>
<td>6</td>
<td>Alpha</td>
<td>Diploma Type Code. From DIPLOMAS table</td>
</tr>
<tr>
<td>44</td>
<td>DIPLOMA_NAME</td>
<td>30</td>
<td>Alpha</td>
<td>Diploma Name. From DIPLOMAS table</td>
</tr>
<tr>
<td>45</td>
<td>MINISTRY_CODE</td>
<td>6</td>
<td>Alpha</td>
<td>Diploma Name From DIPLOMAS table. This is used by CTE as &quot;Course of study&quot;.</td>
</tr>
<tr>
<td>46</td>
<td>REGISTRATION_DATE</td>
<td>Date</td>
<td>Student’s registration date. Date student is &quot;in the seat&quot;.</td>
<td>18-AUG-10</td>
</tr>
<tr>
<td>47</td>
<td>AIG_FLAG</td>
<td>1</td>
<td>Alpha</td>
<td>Academically or intellectually gifted. One character values: B= 01-academically gifted in reading and math – AG R= 20-academically gifted in reading – AR M= 19-academically gifted in math – AM or Null in EXCEPTIONALITY_CODE column of the SPED_EXCEPTIONALITIES table, matched on PUPIL_NUMBER and SCHOOL.</td>
</tr>
<tr>
<td>48</td>
<td>EMAIL_ADDRESS</td>
<td>100</td>
<td>Alpha</td>
<td>Student email address</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

<table>
<thead>
<tr>
<th></th>
<th>CD_PROGRAM_CODE</th>
<th></th>
<th>Alpha</th>
<th>Career development program code - intended career cluster Value descriptions are found in CD_PROGRAM_CODES table.</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>MOD_USER</td>
<td>30</td>
<td>Alpha</td>
<td>User that last modified the record</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>MOD_DATE</td>
<td></td>
<td>Date</td>
<td>Date record was last modified</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>RUN_DATE</td>
<td></td>
<td>Timestamp</td>
<td>Timestamp when record loaded in to repository - YYYYMMDD:HH24:MI:SS</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** Parenthesis indicate NCWISE column name when different from APEX Repository column name.
Appendix J – UID Staff System Extract Data Definitions

Below are sample extracts from UID using MFTS. These give a sample of the type of data fields that will need to be profiled and transformed by the DI-MS.

Purpose: M-F export from the UID Staff System of staff identification data.

Extract Filename: UID All Staff Extract Version 1.1

File Type: 00000

Delimiter: Tab

Last Updated: 5/25/12 5:04 PM

MGID: 1022

Escape: None

Sample Filename: UID1-20111011060044-00-1022-00000-0-ALLSTAFF_V_1_1.txt

Population:

• Includes all staff records

Fields:

<table>
<thead>
<tr>
<th>Col</th>
<th>Field</th>
<th>Size</th>
<th>Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Staff ID</td>
<td>10</td>
<td>Text</td>
<td>The identifier assigned by the eScholar Uniq-ID for Staff system.</td>
<td>1234567890</td>
</tr>
<tr>
<td>2</td>
<td>School Code</td>
<td>6</td>
<td>Text</td>
<td>The staff member’s organizational location.</td>
<td>367</td>
</tr>
<tr>
<td>3</td>
<td>District Code</td>
<td>8</td>
<td>Text</td>
<td>The staff member’s district.</td>
<td>400</td>
</tr>
<tr>
<td>4</td>
<td>Last Name Long</td>
<td>60</td>
<td>Text</td>
<td>The staff member’s legal last name (surname).</td>
<td>Smith</td>
</tr>
<tr>
<td>5</td>
<td>First Name Long</td>
<td>60</td>
<td>Text</td>
<td>This field should be the staff member’s legal first name.</td>
<td>John</td>
</tr>
<tr>
<td>6</td>
<td>Middle Name</td>
<td>60</td>
<td>Text</td>
<td>This field should be the staff member’s legal middle name.  This value can be either the middle initial or the full middle name.</td>
<td>Thomas</td>
</tr>
<tr>
<td>7</td>
<td>Name Suffix</td>
<td>10</td>
<td>Text</td>
<td>The staff member’s suffix for his/her surname, if applicable.</td>
<td>Jr.</td>
</tr>
</tbody>
</table>
## Developing a Data Integration Service

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender Code</td>
<td>Text</td>
<td>The staff member’s gender.</td>
<td>M or F</td>
</tr>
<tr>
<td>Birth Date</td>
<td>Date as Text</td>
<td>The staff member’s date of birth.</td>
<td>10/01/1995</td>
</tr>
<tr>
<td>Alternate Staff ID</td>
<td>Text</td>
<td>The identifier used in the local data system to uniquely identify the staff member. The primary purpose of this field is to provide a mechanism for the local agency to import the Staff ID assigned by the system back into their local systems.</td>
<td>3432439a4</td>
</tr>
<tr>
<td>Social Security Number</td>
<td>Text</td>
<td>The staff member’s social security number.</td>
<td>123455931</td>
</tr>
<tr>
<td>Hispanic Ethnicity Indicator</td>
<td>Text</td>
<td>This field identifies whether or not the staff member is of Hispanic or Latino descent.</td>
<td>0 or 1</td>
</tr>
<tr>
<td>Ethnic Code Short</td>
<td>Text</td>
<td>The new Race and Ethnicity (RE) codes allow staff members to be identified by more than one race and/or ethnicity. The new RE codes more accurately reflect a staff member’s racial and ethnic background by not limiting responses to only one racial or ethnic category, and expand reporting options. Note: The Ethnic Code Short field has been expanded from 2 to 10 characters in length. DPI requires the use of the first 6 characters (left to right) to collect the race/ethnicity details for a staff person. The new RE codes represent the Hispanic Indicator and 5 race codes, each requiring one position within the CEDARS Staff File, for a total length of 6. The position of the Hispanic Indicator and 5 race codes are fixed and are defined as follows: 1st Position: HI7 - Hispanic Indicator (Required; 0=No, 1=Yes) 2nd Position: AM7 - American Indian or Alaskan Native (Required; 0=No, 1=Yes) 3rd Position: AS7 - Asian (Required; 0=No, 1=Yes) 4th Position: BL7 - Black (Required; 0=No, 1=Yes) 5th Position: PI7 - Native Hawaiian or Other Pacific Islander (Required; 0=No, 1=Yes)</td>
<td>100111 (identified as: HI7, BL7, PI7, WH7)</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

<table>
<thead>
<tr>
<th>Column</th>
<th>Label</th>
<th>Length</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Race or Ethnicity Subgroup code</td>
<td>2</td>
<td>The staff member’s racial or ethnic subgroup. For example, if the staff member is American Indian/Alaska Native, this field could represent the specific tribe to which he/she belongs.</td>
<td>Leave blank</td>
</tr>
<tr>
<td>15</td>
<td>Address 1</td>
<td>30</td>
<td>The street name and number of the staff member’s home address.</td>
<td>15 East River Road</td>
</tr>
<tr>
<td>16</td>
<td>Address 2</td>
<td>30</td>
<td>This field should contain additional information concerning the staff member’s home address, such as apartment number.</td>
<td>Suite 102</td>
</tr>
<tr>
<td>17</td>
<td>City</td>
<td>25</td>
<td>The city of the staff member’s home address.</td>
<td>Springfield</td>
</tr>
<tr>
<td>18</td>
<td>State</td>
<td>2</td>
<td>The state of the staff member’s home address.</td>
<td>NY (or “OC” for non-US address)</td>
</tr>
<tr>
<td>19</td>
<td>Full Zip Code</td>
<td>10</td>
<td>The official US postal code for the staff member’s home address.</td>
<td>10023</td>
</tr>
<tr>
<td>20</td>
<td>Job Class Code</td>
<td>2</td>
<td>The staff member’s primary position within the location and/or district.</td>
<td>Leave blank</td>
</tr>
<tr>
<td>21</td>
<td>Original Hire Date</td>
<td>10</td>
<td>The date the location/district originally hired the staff member.</td>
<td>10/10/2006</td>
</tr>
<tr>
<td>22</td>
<td>Highest Degree Earned</td>
<td>20</td>
<td>The type of degree or certificate the staff member has earned. If the staff member has multiple degrees/certificates, the highest level should be used.</td>
<td>Leave blank</td>
</tr>
<tr>
<td>23</td>
<td>Annual Salary</td>
<td>10</td>
<td>The total compensation for the staff member for services provided to the location/district. This number should include any additional stipends the staff member receives from the location/district for such services as coaching athletic</td>
<td>45000</td>
</tr>
</tbody>
</table>
Developing a Data Integration Service

teams or moderating after school activities, but still within the context of the primary designation.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>Full Staff Name</td>
<td>255</td>
<td>Text</td>
<td>The staff member’s full name, using the form Last Name, First Name (e.g., Doe, Jane). The middle name can also be included. This value will be primarily used for reporting and/or other user interface display purposes.</td>
</tr>
<tr>
<td>25</td>
<td>Itinerant Teacher</td>
<td>3</td>
<td>Text</td>
<td>Indicates whether or not the staff member has been classified as an Itinerant Teacher (on loan from another agency or a staff member who provides services in more than one agency).</td>
</tr>
<tr>
<td>26</td>
<td>Active/Inactive Indicator</td>
<td>1</td>
<td>Text</td>
<td>The staff member’s current status (e.g., Active/Inactive) in the location/district.</td>
</tr>
<tr>
<td>27</td>
<td>Previous Last Name</td>
<td>60</td>
<td>Text</td>
<td>The former last name of the staff member, if applicable. The Previous Last Name data can include changes in last name due to marriage, divorce, legal name change or other forms of name changes.</td>
</tr>
<tr>
<td>28</td>
<td>Last Update</td>
<td>10</td>
<td>Date as Text</td>
<td>The date of the last update for this staff record.</td>
</tr>
</tbody>
</table>
## Developing a Data Integration Service

### Appendix K – DPI Authoritative Data Source List

<table>
<thead>
<tr>
<th>Line</th>
<th>Data Grouping</th>
<th>Data Type</th>
<th>Critical Dates</th>
<th>Program Contact</th>
<th>Business Leadership</th>
<th>IT Leadership</th>
<th>Authoritative Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accountability</td>
<td>All data specific to student accountability and assessment</td>
<td>First day of Fall Testing and First Day of Spring Testing</td>
<td>Tammy Howard</td>
<td>Angela Quick</td>
<td>Ken Barbour</td>
<td>Accountability</td>
</tr>
<tr>
<td>2</td>
<td>Alternative Learning Programs</td>
<td>All data specific to alternative learning programs.</td>
<td>Annual roster data requested by July 1; Data collection closed Sept. 1</td>
<td>Ken Gattis</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>NC WISE</td>
</tr>
<tr>
<td>3</td>
<td>Budgetary Allotments</td>
<td>LEA Budgetary data</td>
<td>Bi-weekly</td>
<td>Alexis Schauss</td>
<td>Alexis Schauss</td>
<td>Donna Roch</td>
<td>BUD</td>
</tr>
<tr>
<td>4</td>
<td>Career and Technical Education</td>
<td>All data specific to the CTE program</td>
<td>December 31</td>
<td>Rhonda Welfare</td>
<td>Maria Pitre-Martin</td>
<td>VACANT</td>
<td>CTE System</td>
</tr>
<tr>
<td>5</td>
<td>Course Codes</td>
<td>Data specific to course codes</td>
<td>As courses are added.</td>
<td>Maria Pitre-Martin</td>
<td>Angela Quick</td>
<td>George Batten</td>
<td>NC WISE</td>
</tr>
<tr>
<td>6</td>
<td>Crime and Violence Data</td>
<td>All data specific to acts of crime and violence</td>
<td>Annual data verification period: June 30 - August 30</td>
<td>Ken Gattis</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>NC WISE</td>
</tr>
<tr>
<td>7</td>
<td>Dropout</td>
<td>All data specific to the Dropout program. State and Federal %’s are calculated using different business rules.</td>
<td>“Effective date” is during fourth week of September. Collection and verification ends October 1.</td>
<td>Ken Gattis</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>Dropout Collection</td>
</tr>
<tr>
<td>8</td>
<td>Economically Disadvantaged Students</td>
<td>All data specific to the free and reduced lunch program</td>
<td>Extractions: December 1 and April 1</td>
<td>Lynn Harvey</td>
<td>Ben Matthews</td>
<td>Jeff Merritt</td>
<td>Child Nutrition System</td>
</tr>
<tr>
<td>9</td>
<td>Exceptional Children</td>
<td>All data specific to the Exceptional Children Program</td>
<td>Active Child Count: April 1 and December 1; Exit Child Count: September 1. Data Available one month after.</td>
<td>Ira Wolfe</td>
<td>Mary Watson</td>
<td>Quentin Parker</td>
<td>CECAS</td>
</tr>
<tr>
<td>10</td>
<td>Financial External</td>
<td>LEA financial data</td>
<td>Monthly</td>
<td>Alexis Schauss</td>
<td>Alexis Schauss</td>
<td>Donna Roch</td>
<td>DBS External</td>
</tr>
<tr>
<td>11</td>
<td>Financial Internal</td>
<td>Agency financial data</td>
<td>CAFR Closeout: August 31</td>
<td>Peta-Gaye Shaw</td>
<td>Sarah Harris</td>
<td>Donna Roch</td>
<td>DBS Internal</td>
</tr>
<tr>
<td>12</td>
<td>Homeless</td>
<td>All data specific to the Homeless Program</td>
<td>Collection Period: August-May 2011; Data Available two weeks after.</td>
<td>Loreto Tessini</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>CFDC-Homeless</td>
</tr>
</tbody>
</table>

DPI Data Management Group (DMG)  
May 26, 2010
# Developing a Data Integration Service

## DPI Authoritative Data Source List

<table>
<thead>
<tr>
<th>Line</th>
<th>Data Grouping</th>
<th>Data Type</th>
<th>Critical Dates</th>
<th>Program Contact</th>
<th>Business Leadership</th>
<th>IT Leadership</th>
<th>Authoritative Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Immigrant</td>
<td>All data specific to the Immigrant program</td>
<td>Collection Pull: February 2 &amp; 3</td>
<td>Glenda Harrell</td>
<td>Helga Fasciano</td>
<td>VACANT</td>
<td>Immigrant Collection</td>
</tr>
<tr>
<td>14</td>
<td>Limited English Proficient</td>
<td>All data specific to the LEP Program</td>
<td>October 1 Headcount Collection Pull Oct. 2 &amp; 3; Verification: May; Data Before YET Completed: June; Testing Label: late November.</td>
<td>Glenda Harrell</td>
<td>Helga Fasciano</td>
<td>VACANT</td>
<td>LEP Collection</td>
</tr>
<tr>
<td>15</td>
<td>Migrant</td>
<td>All data specific to the Migrant Program</td>
<td>Certified 2010-11 Data Available 10/1/11</td>
<td>Loreto Tessini</td>
<td>Donna Brown</td>
<td>Vendor</td>
<td>MIS2000</td>
</tr>
<tr>
<td>16</td>
<td>Neglected and Delinquent</td>
<td>All data specific to the Neglected and Delinquent Program</td>
<td>Collection Period: August-May 2011; Data Available two weeks after.</td>
<td>Loreto Tessini</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>CFDC-N&amp;D</td>
</tr>
<tr>
<td>17</td>
<td>School Directory Information</td>
<td>All school contact information data (school name/code, LEA name/code, address, phone, email, website, chief school officers)</td>
<td>Updates: Before Education Directory Printed and Before School Report Cards Preview.</td>
<td>Nicola Lefler</td>
<td>Alexis Schauss</td>
<td>Donna Roch</td>
<td>School Name and Address System</td>
</tr>
<tr>
<td>18</td>
<td>Student Information</td>
<td>Data specific to students</td>
<td>Daily</td>
<td>Alexis Schauss</td>
<td>Alexis Schauss</td>
<td>George Batten</td>
<td>NC WISE</td>
</tr>
<tr>
<td>19</td>
<td>Supennial Services and School Choice</td>
<td>All data specific to supplemental educational services and school choice</td>
<td>Collection Period: August-May 2011; Data Available two weeks after.</td>
<td>Loreto Tessini</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>CFDC-SES &amp; Choice</td>
</tr>
<tr>
<td>20</td>
<td>Suspensions and Expulsions</td>
<td>All data specific to suspensions and expulsions</td>
<td>Annual data verification period: June 30 - August 30</td>
<td>Ken Gattis</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>NC WISE</td>
</tr>
</tbody>
</table>
## Developing a Data Integration Service

### DPI Authoritative Data Source List

<table>
<thead>
<tr>
<th>Line</th>
<th>Data Grouping</th>
<th>Data Type</th>
<th>Critical Dates</th>
<th>Program Contact</th>
<th>Business Leadership</th>
<th>IT Leadership</th>
<th>Authoritative Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Title I Schools</td>
<td>All data specific to Title I schools and Targeted Assistance and Schoolwide Programs</td>
<td>Collection Period: TAS: August-May 2011; Title I Status: June 2011; LEA &amp; SCH Improvement Status: September 2011; Data Available two weeks after</td>
<td>Loreto Tessini</td>
<td>Donna Brown</td>
<td>VACANT</td>
<td>CFDC-TAS, CFDC-ESSR, CFDC-LEA Improvement, CFDC-School Improvement</td>
</tr>
<tr>
<td>22</td>
<td>Technology data</td>
<td>All data specific to LEA/SCH Technology</td>
<td>AMTR Data harvested bi-annually: January &amp; June/July</td>
<td>Melanie Honeycut</td>
<td>Neill Kimrey</td>
<td>VACANT</td>
<td>AMTR</td>
</tr>
<tr>
<td>23</td>
<td>Graduation Data Verification System</td>
<td>All data specific to students graduation</td>
<td>June 30th</td>
<td>Alexis Schauss</td>
<td>Alexis Schauss</td>
<td>Donna Roch</td>
<td>Graduate Survey</td>
</tr>
</tbody>
</table>
## Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPI</td>
<td>North Carolina Department of Public Instruction[^31]</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract, Transform, Load</td>
</tr>
<tr>
<td>UID</td>
<td>eScholar Uniq-ID[^32]</td>
</tr>
<tr>
<td>IAM</td>
<td>Identity and Access Management</td>
</tr>
<tr>
<td>LEA</td>
<td>Local Education Agency (school district)</td>
</tr>
<tr>
<td>MFTS</td>
<td>Managed File Transfer System</td>
</tr>
<tr>
<td>NCID</td>
<td>North Carolina Identity Management System[^33]</td>
</tr>
<tr>
<td>NCWISE</td>
<td>North Carolina Window on Student Education[^34]</td>
</tr>
<tr>
<td>ODS</td>
<td>Operational Data Store</td>
</tr>
<tr>
<td>RFI</td>
<td>Request for Information</td>
</tr>
<tr>
<td>ITS</td>
<td>North Carolina Information Technology Services</td>
</tr>
<tr>
<td>DI-MS</td>
<td>Data Integration Managed Service</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposals</td>
</tr>
<tr>
<td>RttT</td>
<td>Race to the Top[^35]</td>
</tr>
</tbody>
</table>

[^33]: [https://ncid.nc.gov](https://ncid.nc.gov)
[^34]: [http://www.ncwise.org/](http://www.ncwise.org/)
This page is left blank intentionally