



FEDERATED & CENTRALIZED MODELS

Tuesday, October 30, 2012

Facilitator: Jim Campbell (SST), Jeff Sellers (SST), Keith Brown (SST)

Panelists:

Charles McGrew, Kentucky P-20 Data Collaborative

Mimmo Parisi, National Strategic Planning & Analysis Research Center (nSPARC)

Neal Gibson, Arkansas Research Center

Aaron Schroeder, Virginia Tech



AGENDA

- Background/Overview
- Rationale for choosing a model
- Infrastructure and design
- Data Access
- Additional information



BACKGROUND/OVERVIEW



KENTUCKY

Model:

Centralized data warehouse that brings the data into a neutral location (third-party) where they are linked together then de-identified so no agency has access to any other agency's identifiable data.



VIRGINIA

Model:

Virginia is a case study in the difficulties of combining data from multiple agencies while remaining in compliance with federal and state-level privacy requirements

- Traditional Data Integration Issues
- Public Sector Specific Integration Issues
- Virginia Specific Issues



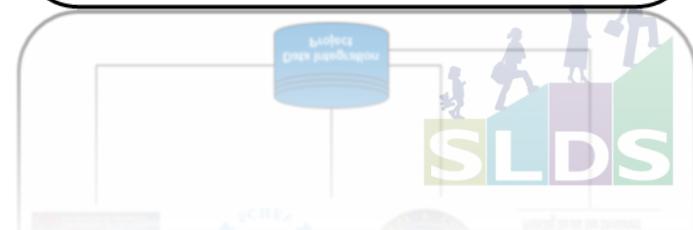
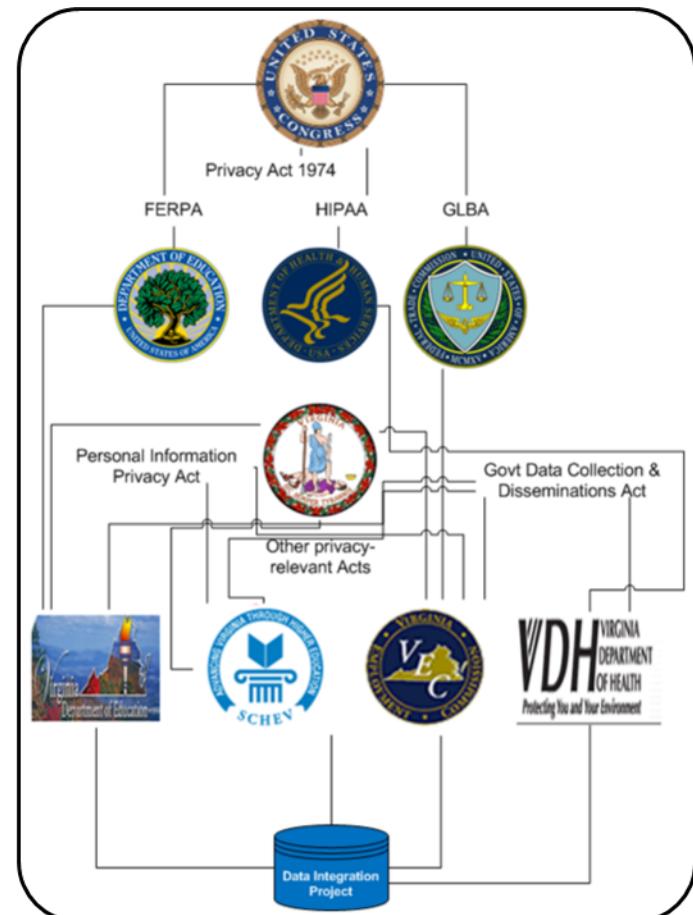
VIRGINIA

Implementation Environment Public Sector Statutory and Regulatory Heterogeneity

Multiple levels of statutory law

Multiple implementations of
regulatory law at each level of
statutory law

Most conservative interpretation of
regulatory law becomes de facto
standard



ARKANSAS

Arkansas Research Center



File A



File B



Your knowledge is limited to what's
in these two files ONLY



ARKANSAS

Knowledge Base Approach:

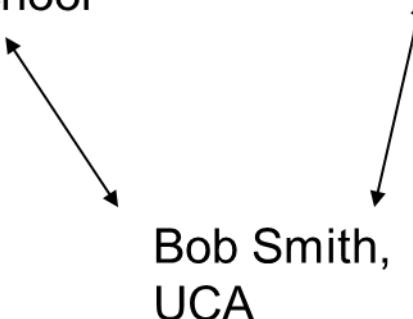
All known representations are stored to facilitate matching in the future and possibly resolve past matching errors.



Bob Smith,
Conway
High School



Robert Smith,
Acxiom



Knowledge Base

Cluster	Representation
KB5765	Bob Smith, CHS
KB5765	Robert Smith, Acxiom
KB5765	Bob Smith, UCA



MISSISSIPPI



STATEWIDE LONGITUDINAL DATA SYSTEM (SLDS)

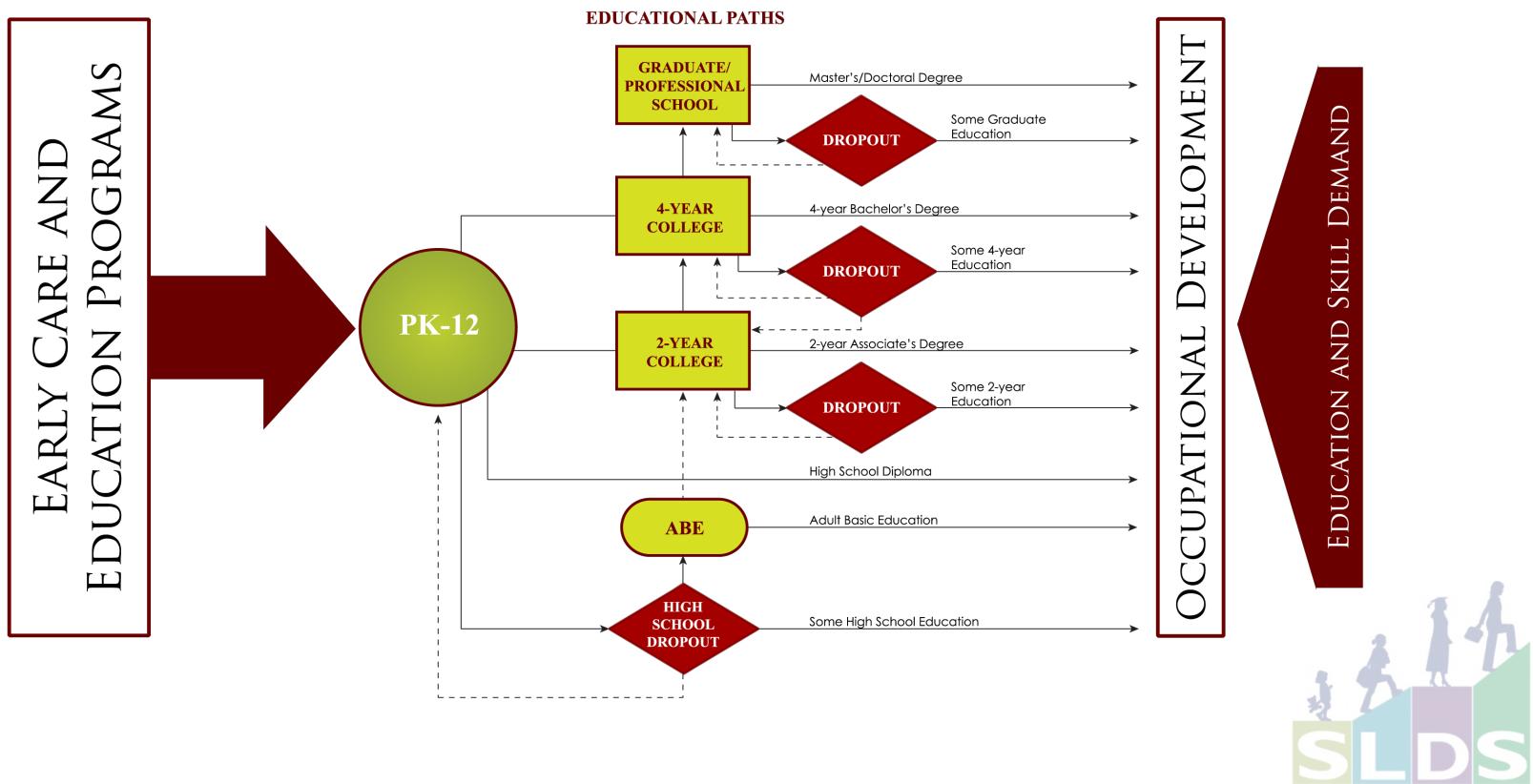


MISSISSIPPI



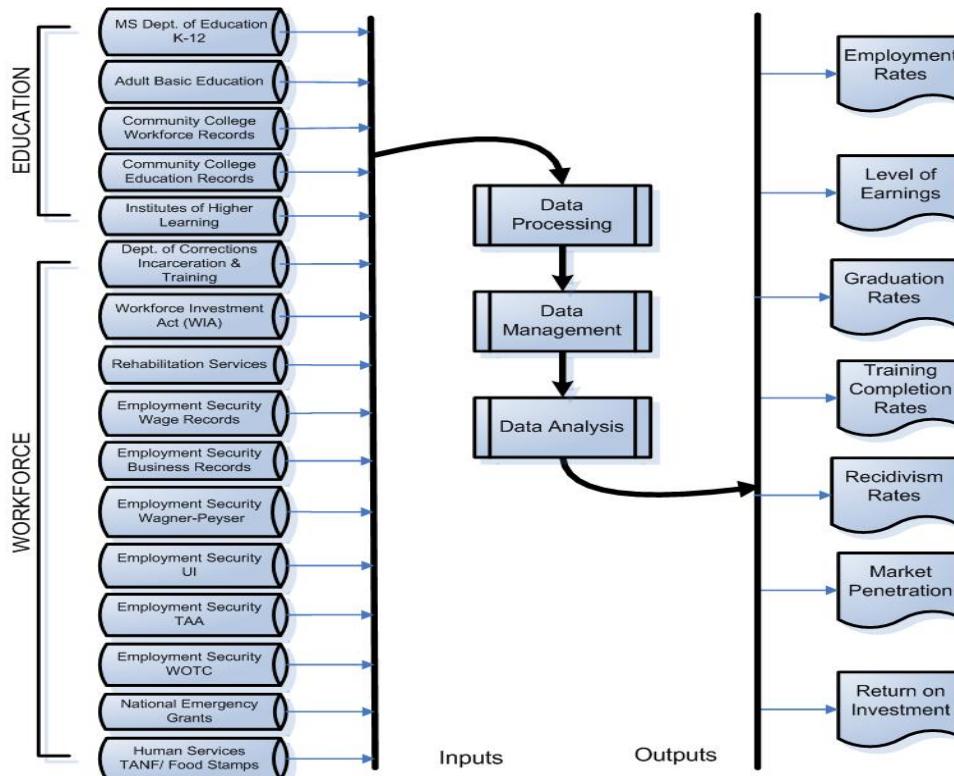
MISSISSIPPI

SLDS CONCEPTUAL MODEL



MISSISSIPPI

DATA WAREHOUSE MODEL



RATIONALE FOR CHOOSING A MODEL



KENTUCKY

- Not all agencies we want to participate have data warehouses and could participate in a federated model.
- Upgrades and infrastructure changes in participating agencies would interrupt “service”.
- Political concerns about changing leadership that may allow agencies to simply stop participating.
- Centralizing data allows for shared resources and cost savings over more “silo” data systems.
- Most agencies lack research staff to utilize the system and there was a desire to centralize some analyses.
- The de-identified model satisfies agency legal concerns.
- The ability to reproduce numbers over time.



VIRGINIA

Implementation Environment and Virginia Specific Limitations:

Structural

- Decentralized authority structure in potential partner agencies (e.g. health, social services) resulting in different data systems, standards, and data collected

Legal

- VA § 2.2-3800: Government Data Collection and Dissemination Practices Act
- VA § 59.1-443.2 - Restricted use of social security numbers
- Assistant Attorneys General interpretations
 - “No one person , inside or outside a government agency, should be able to create a set of identified linked data records between partner agencies”



VIRGINIA

Consolidated Data Systems (Warehouse)

- Can be very expensive (to both build and maintain)
- Too difficult to embody (program) the multiple levels of federal and state statutory and regulatory privacy requirements – must have laws in place to allow for centralized collection
- Lack of clear data authority, per data system, between state agencies and between state and local-level agencies – participation is not compulsory

Federated Data Systems

- System that interacts with multiple data sources on the back-end and presents itself as a single data source on the front-end
- The key to linking up the different data sources is a central linking apparatus
- Allows for the maintenance of existing privacy protection rules and regulations
- Can significantly reduce application development time and cost



FERPA v. 1 Compliance

Match rates are important, especially for multi-agency data

Caleb Gibson



MISSISSIPPI

RATIONALE

- MINIMIZE BURDEN ON DATA STAKEHOLDERS
- MINIMIZE SYSTEM CHANGES



INFRASTRUCTURE AND DESIGN



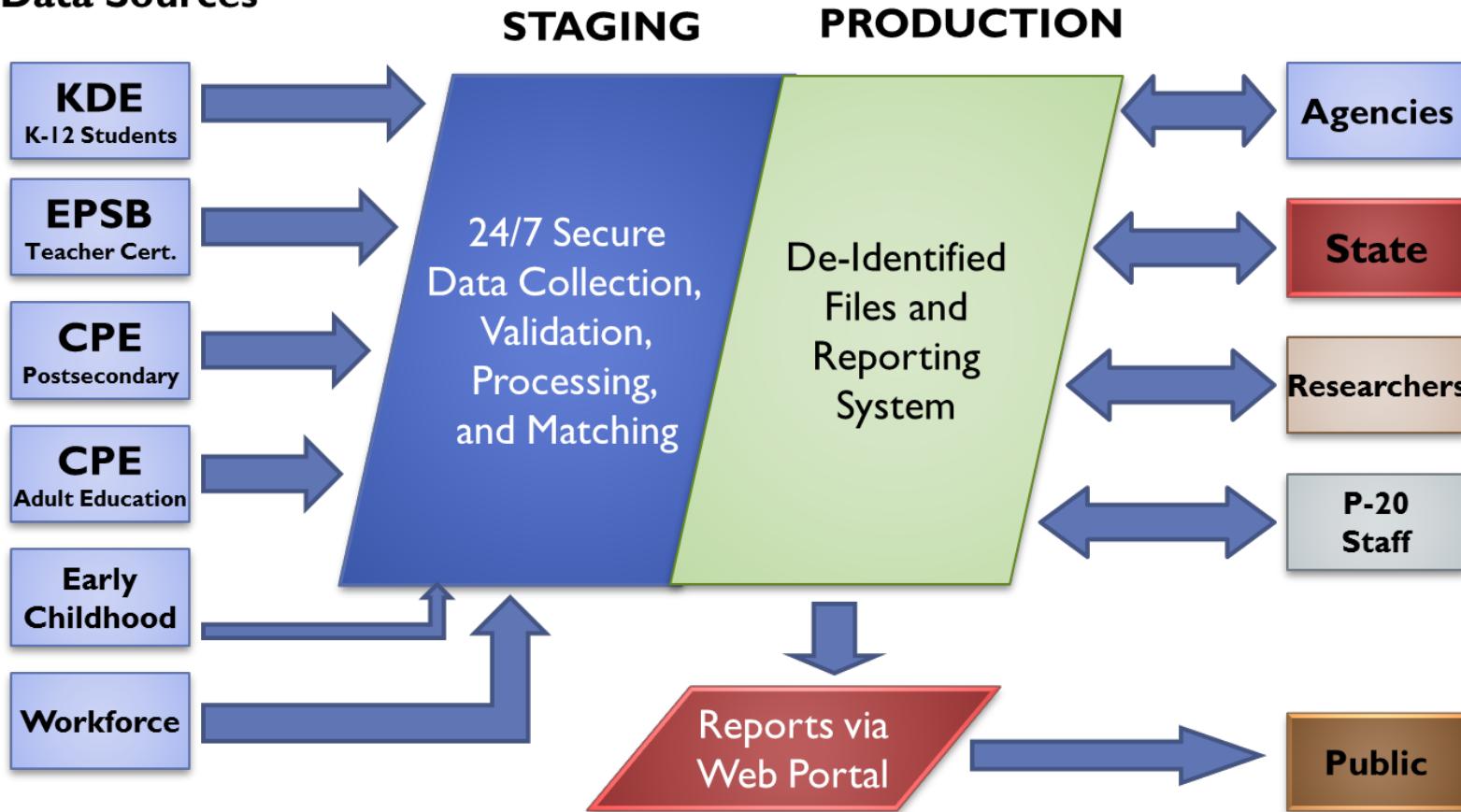
KENTUCKY

- Agencies provide data on a regular schedule and have access to the de-identified system through a standard reporting tool.
- “Master Person” record matching process that becomes “better” over time by retaining all the different versions of data for matching.
- Staging environment where data are validated and checked.

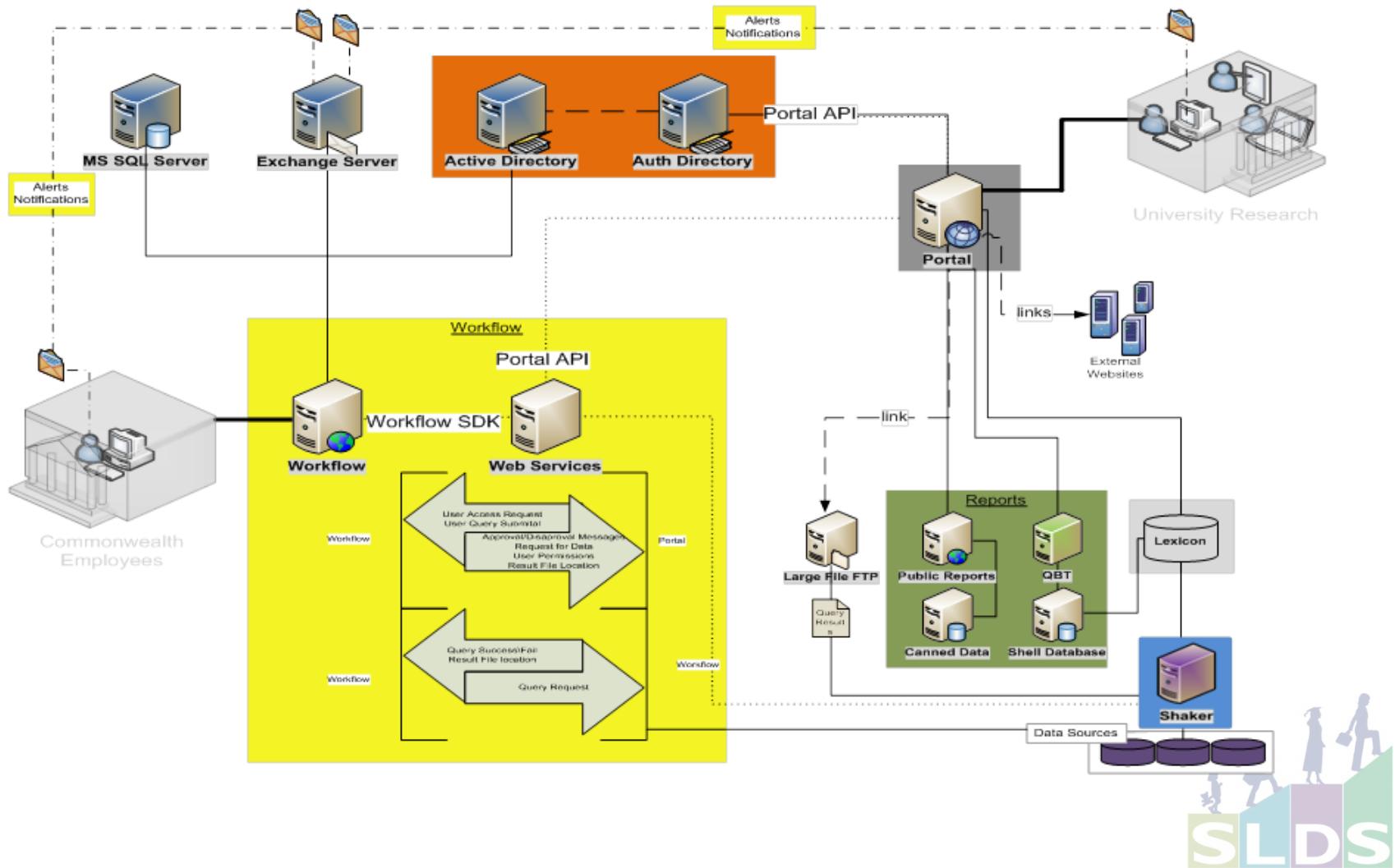


KENTUCKY

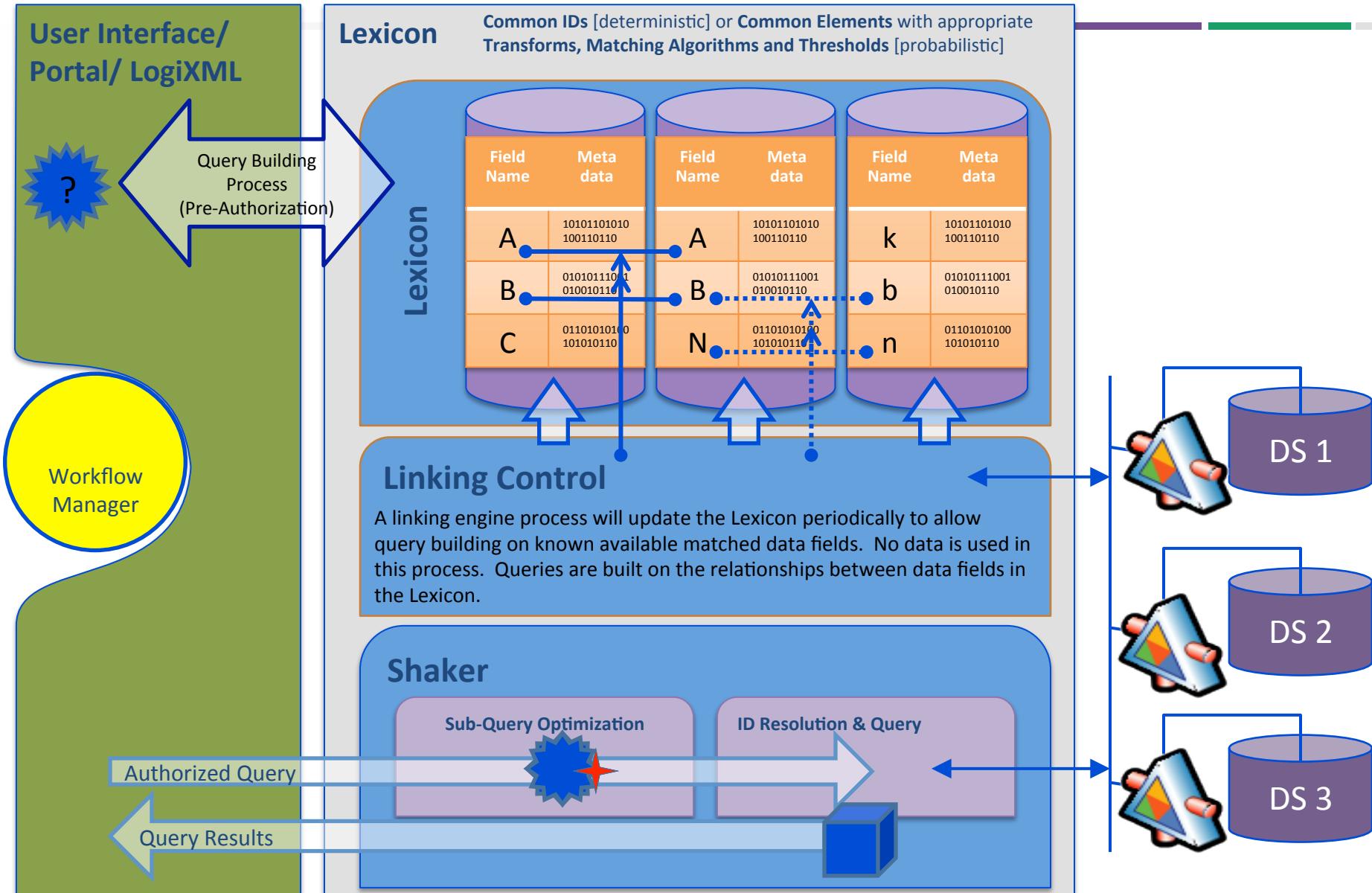
Data Sources



VIRGINIA - WORKFLOW



LEXICON – SHAKER PROCESS OVERVIEW

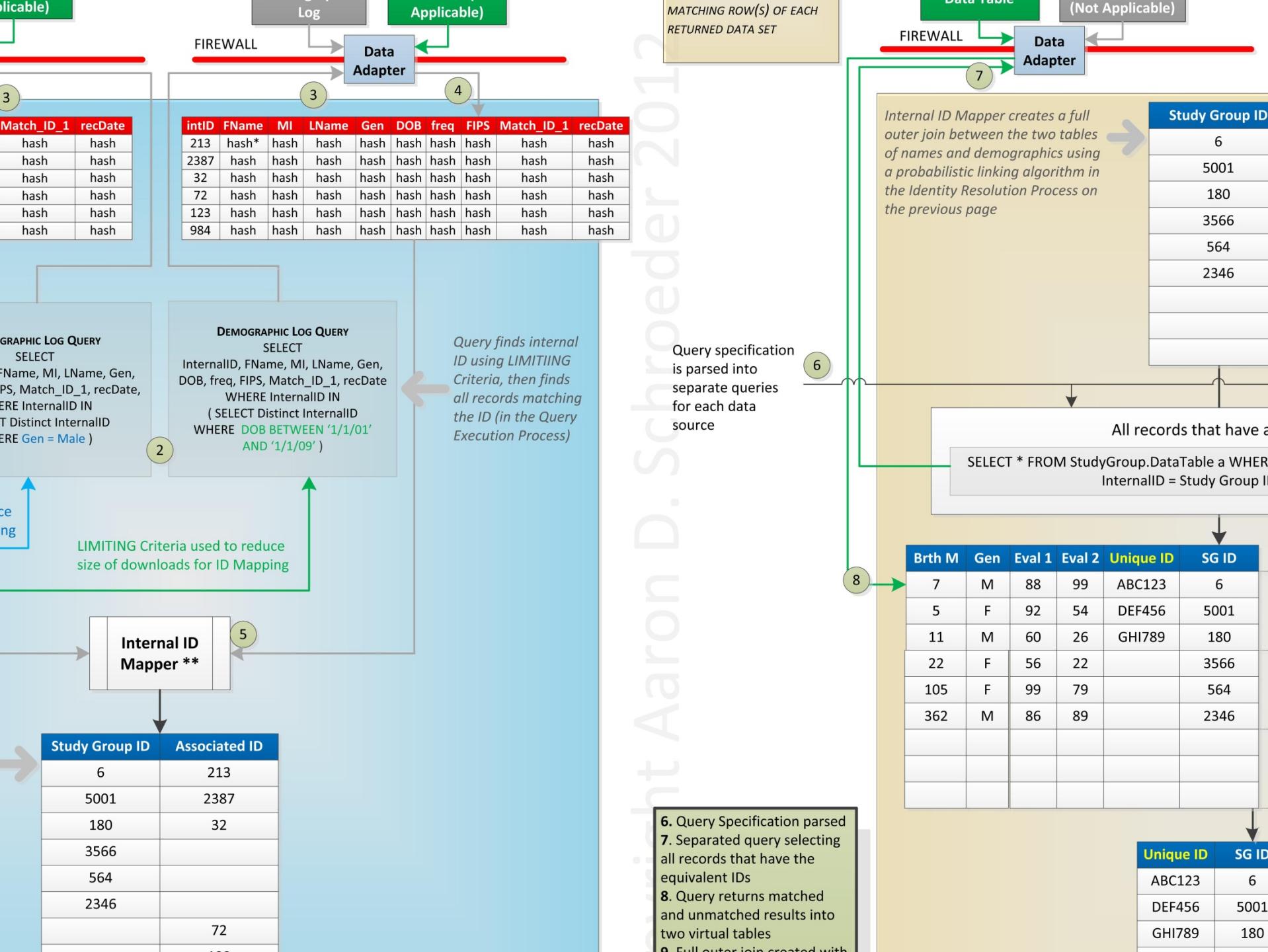


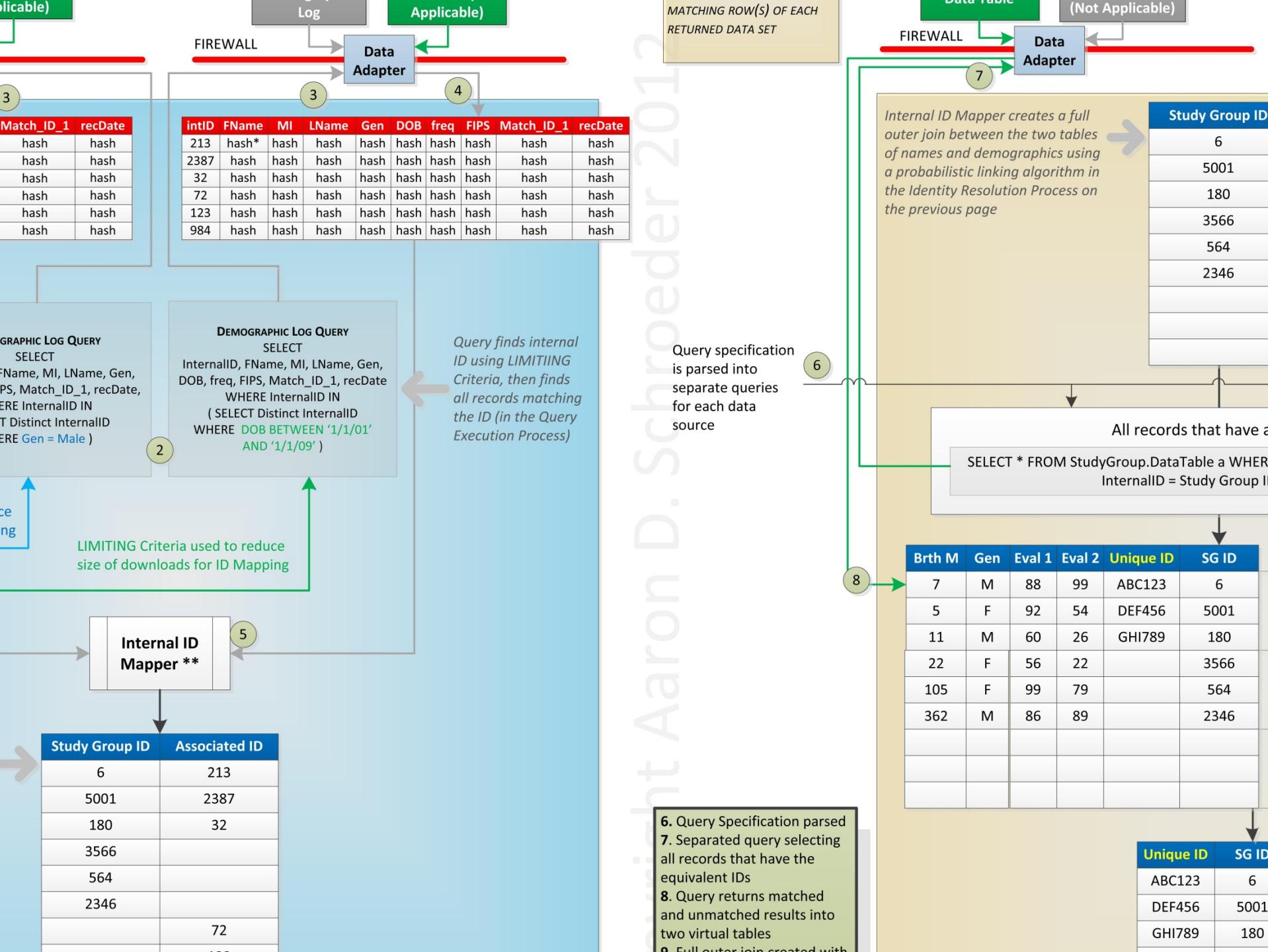
PRIVACY PROTECTING FEDERATED QUERY

Two Steps:

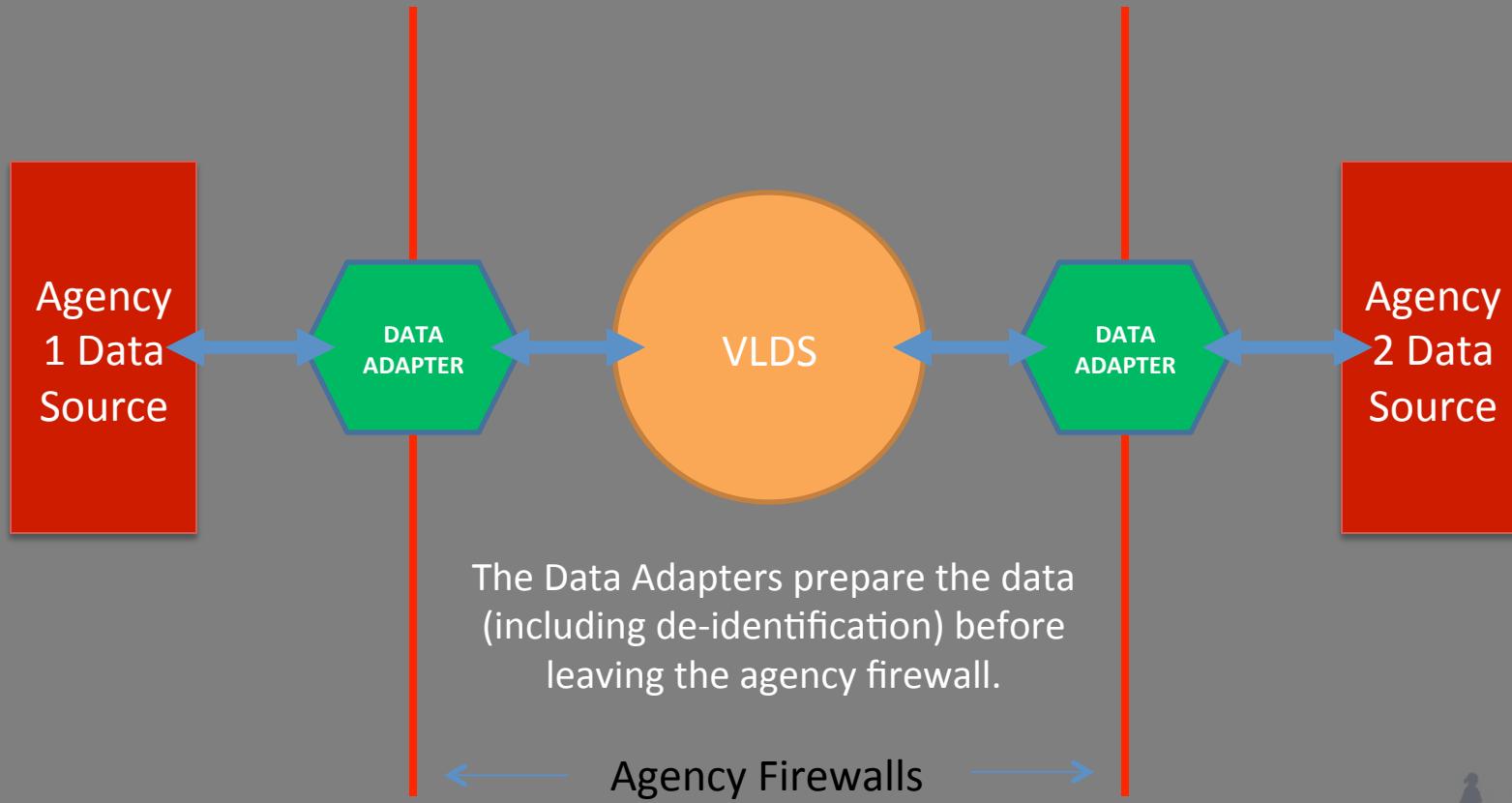
1. Identity Resolution Process
2. Query Execution Process







How De-Identification Works in VLDS System



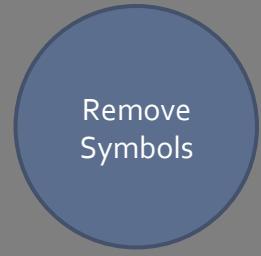
GETTING DATA READY FOR “DE-IDENTIFIED FEDERATION”

What the Data Adapter Does:

De'Smith-Barney IV



De'Smith-Barney



DeSmithBarney

DeSmithBarney

Substitution Cipher

YDXWKQTAGOLCNSVEFHMRJPBZUI



Hashing Key – Dynamically Generated and sent from HANDS

b164f11d-aa37-44ca-93c3-82d3e0155061

C57S78XCEBF9WECP2AA9DK59N1CO27QBES54HFD



Cleaned and Encoded Matching Data (Internal ID, First and Last Name)

INTERNAL_ID_HASHED	FIRST_NAME	LAST_NAME
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	J11EDAV3E
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	J11EDAV3E
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	J11EDAV3E
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	X11SDK3EF86
044AA90CE74E2ED3B6B0B0CFE93F8ED263B73050	F11BDAE3EM86AE8	J11EDAV3E

INTERNAL_ID is the same

LAST_NAME is NOT

Many agencies DO NOT have an Index of unique individuals. There can be many representations of that individual.

What do we do?

Statistical Log
Analysis and
Reduction

We dynamically build a new
“virtual” record made up of
“most likely” demographics



Probabilistic Linkage Process (Creating a Linking Directory)

(After we have a unique person index for each agency dataset)

Blocking

m and u Parameter Calculation

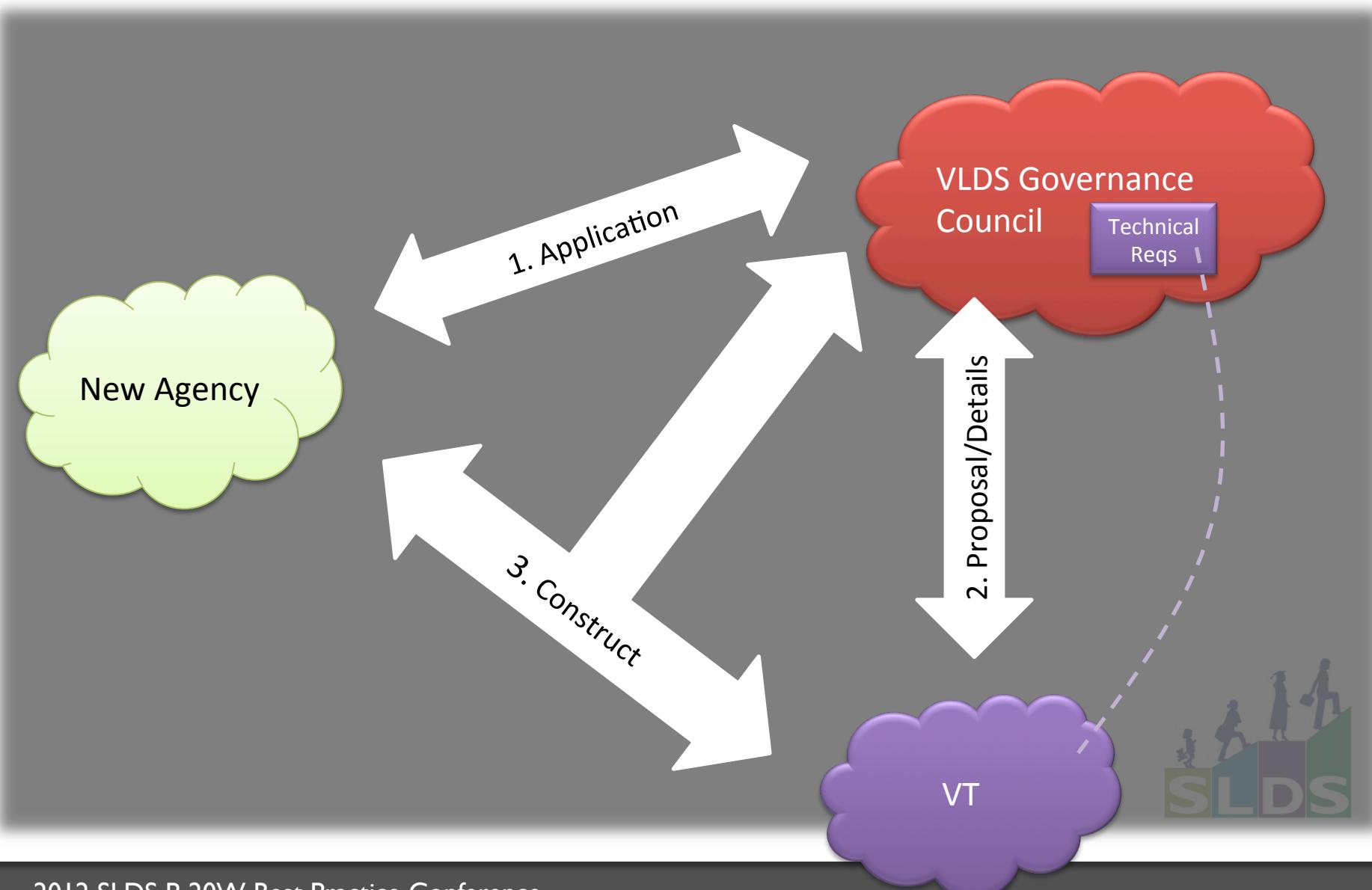
Matching-Column Weight Calculations

Match Scoring

Linkage Determination and addition to Linking Directory

- Linkage Determination – A Cutoff score needs to be set for each blocked comparison, below which a link is not accepted as a real “link”
- The best method of establishing this cutoff is for the system operator to work with a content-area expert to determine the peculiarities of data for that content-area
- In some data sets it may be very unlikely that a birthdate was entered incorrectly, while in another, it may happen very regularly – a computer can not automatically know this
- Once these cutoffs are set, they don’t need to be changed unless something drastic occurs to change the nature of the dataset

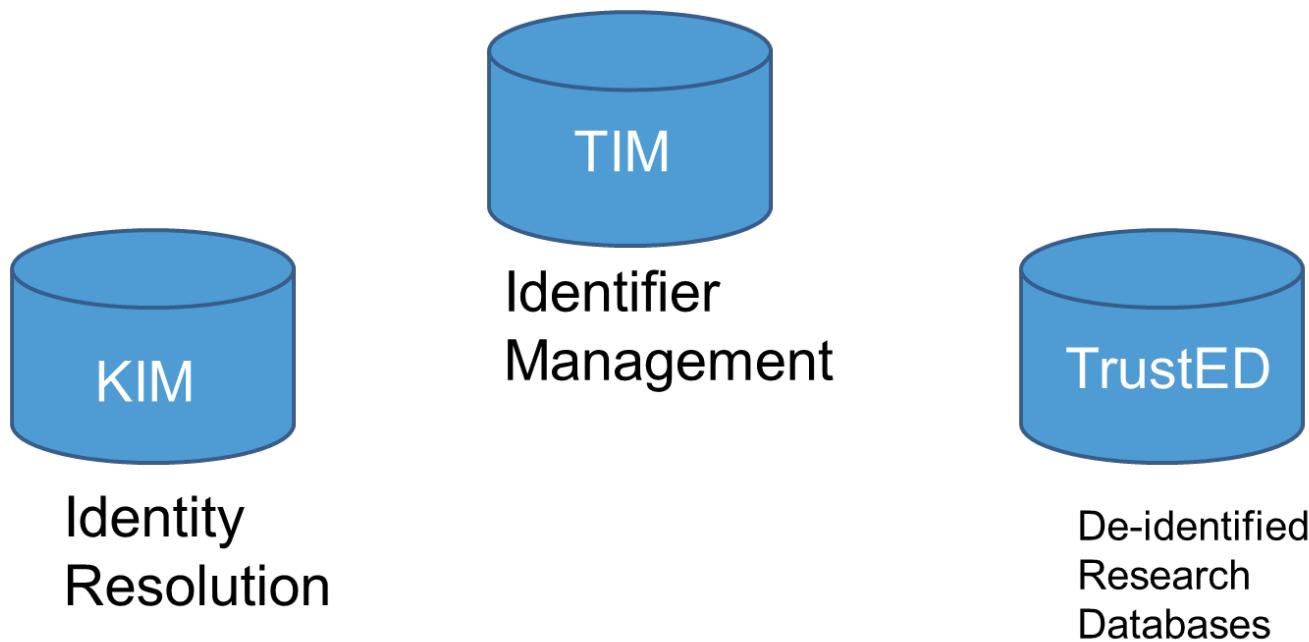
Data Partner Intake Process



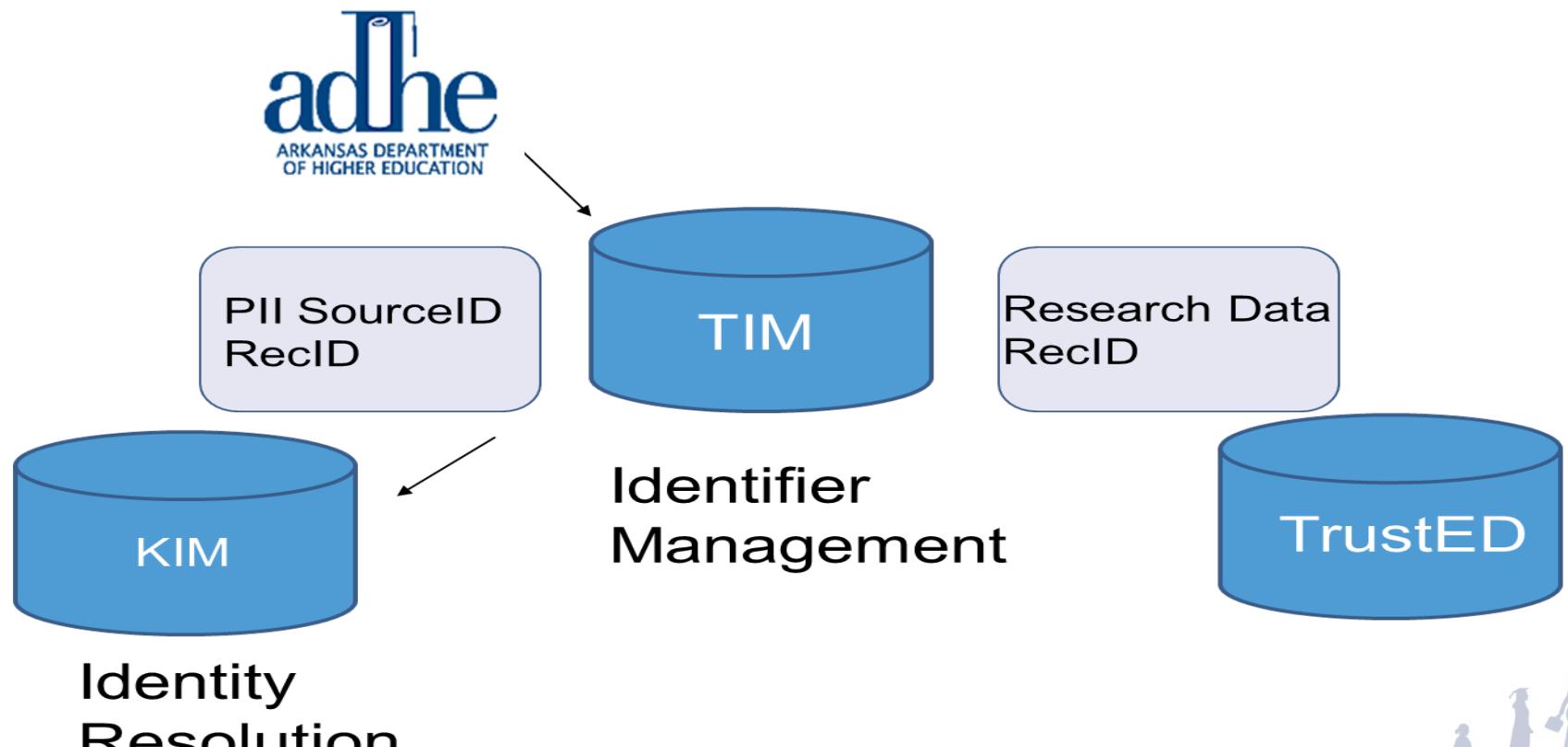
TrustEd:

Knowledgebase Identity Management (KIM)

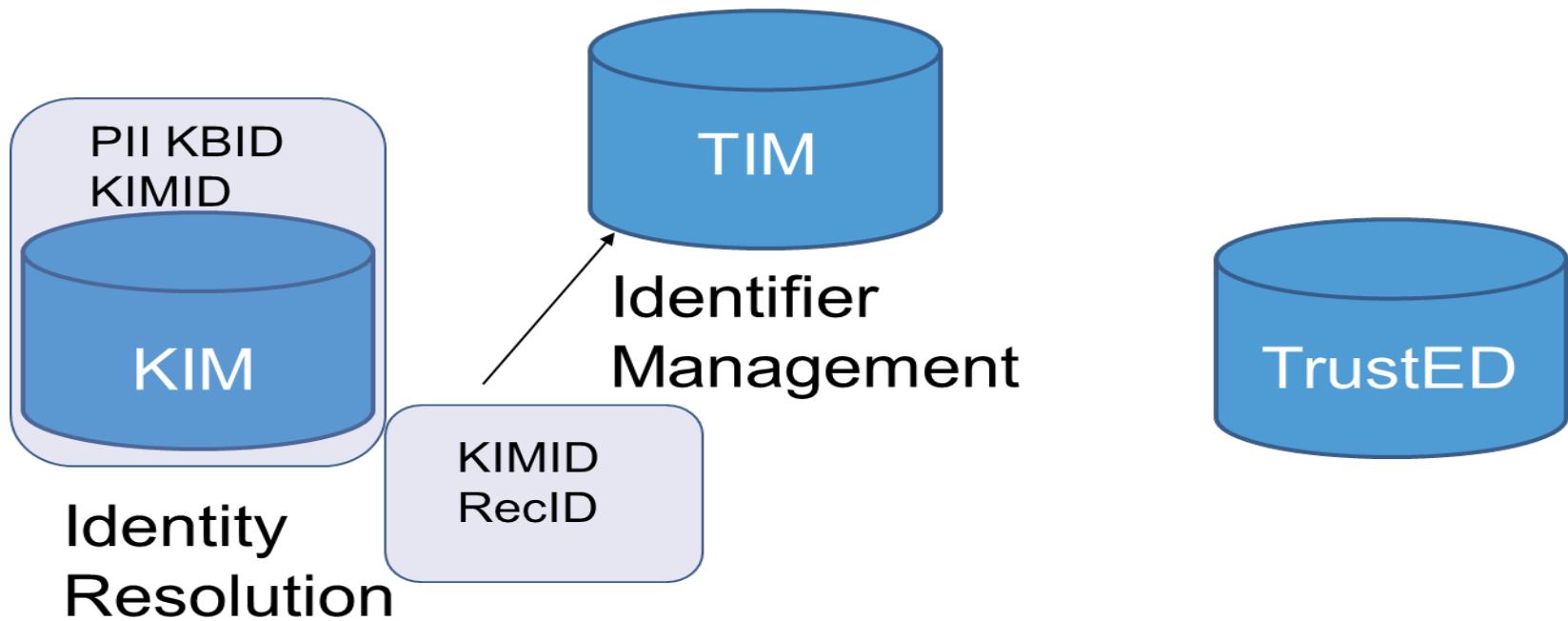
TrustEd Identifier Management (TIM)



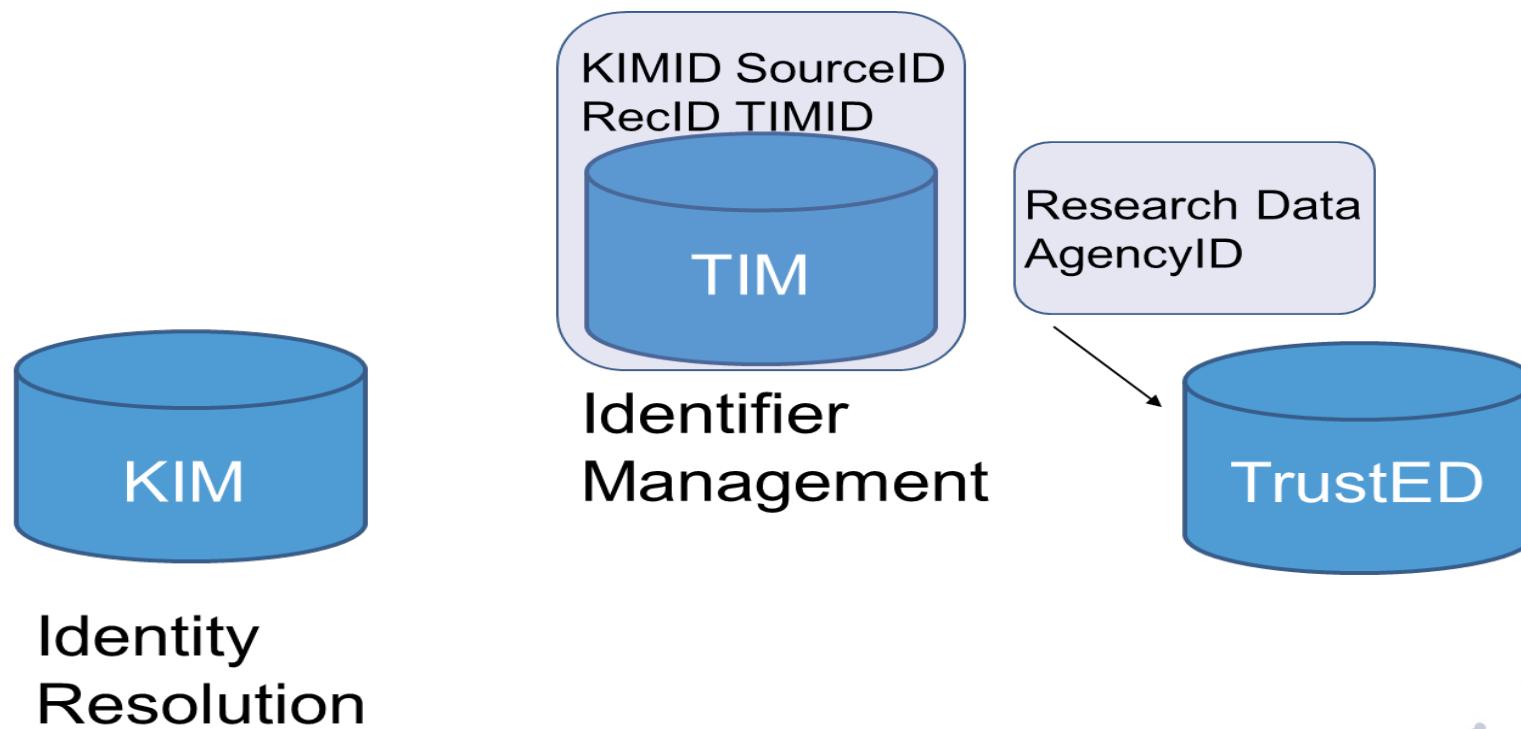
TrustEd: KIM & TIM



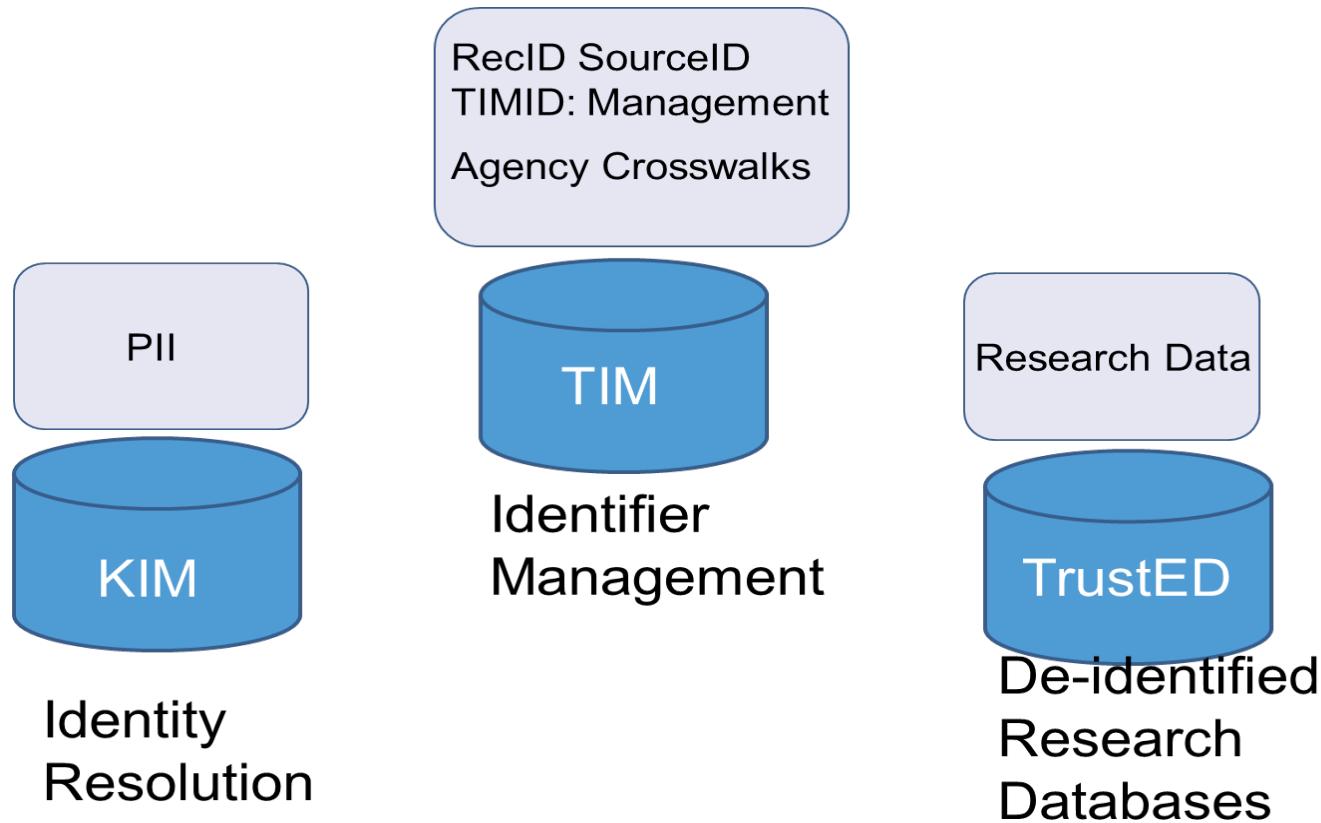
TrustEd: KIM & TIM



TrustEd: KIM & TIM



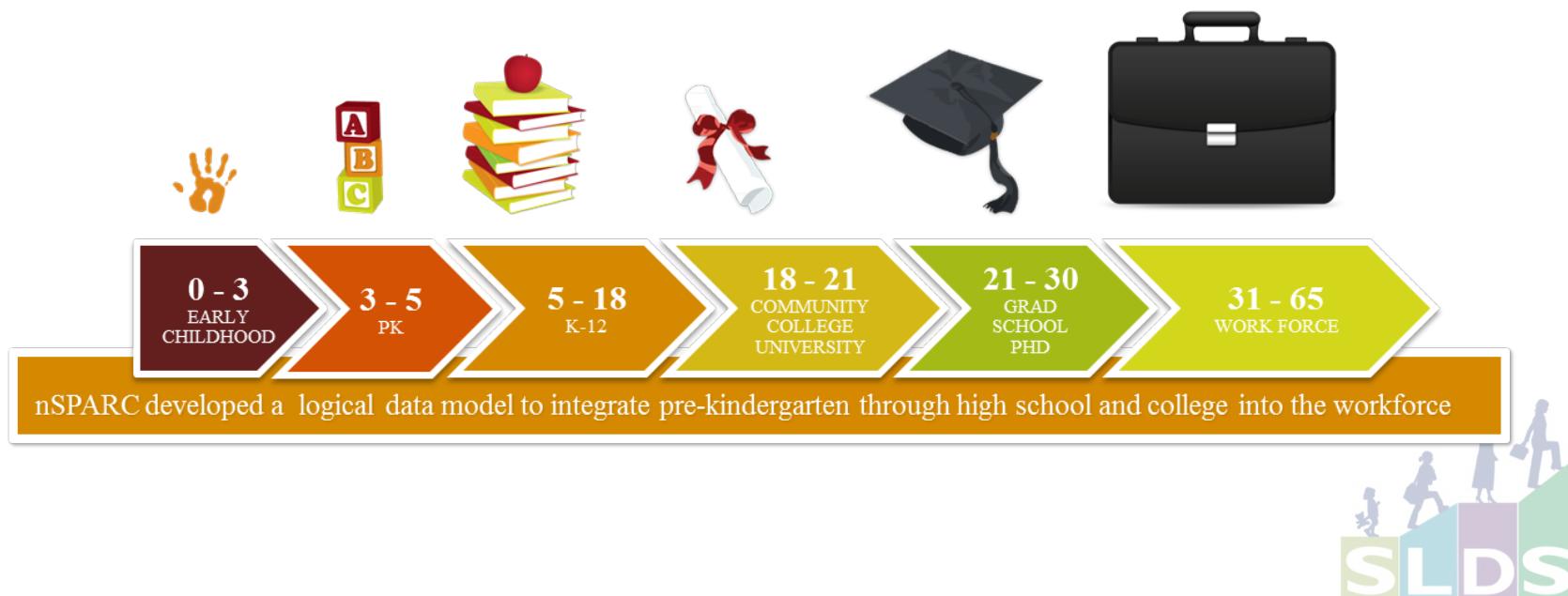
TrustEd: KIM & TIM



MISSISSIPPI

LOGICAL DATA MODEL

The SLDS is designed to put relevant and timely information for better decision making into the hands of appropriate stakeholders (parents, teachers, principals, superintendents, political leaders). It does so by creating an integrated data model and by framing information within the context of individuals, programs, and organizations.



MISSISSIPPI

LOGICAL DATA MODEL

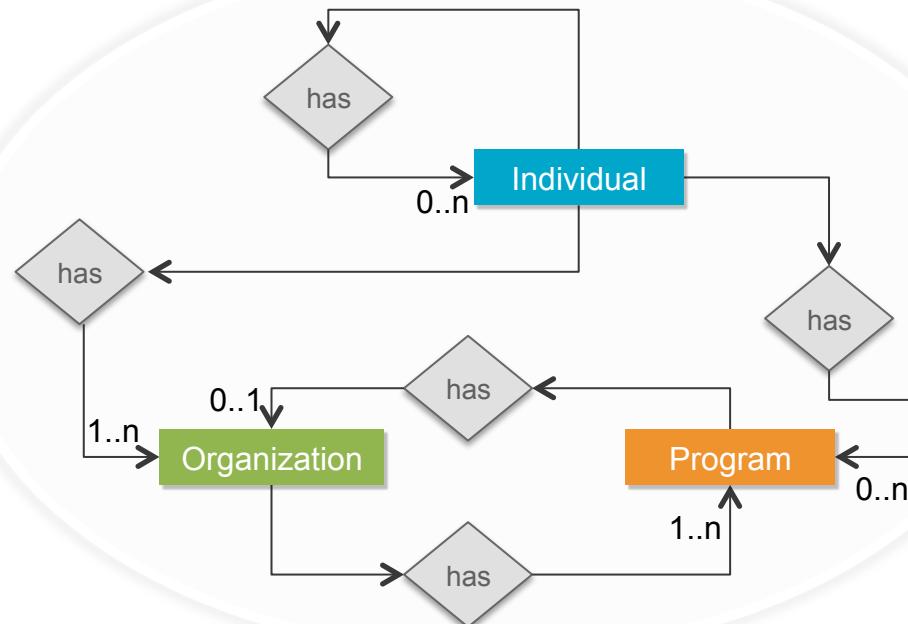
High level overview of data interaction

ORGANIZATION:

- Institution
- Pre School
- School
- Community College
- WIA Local Area
- DOL Service Center

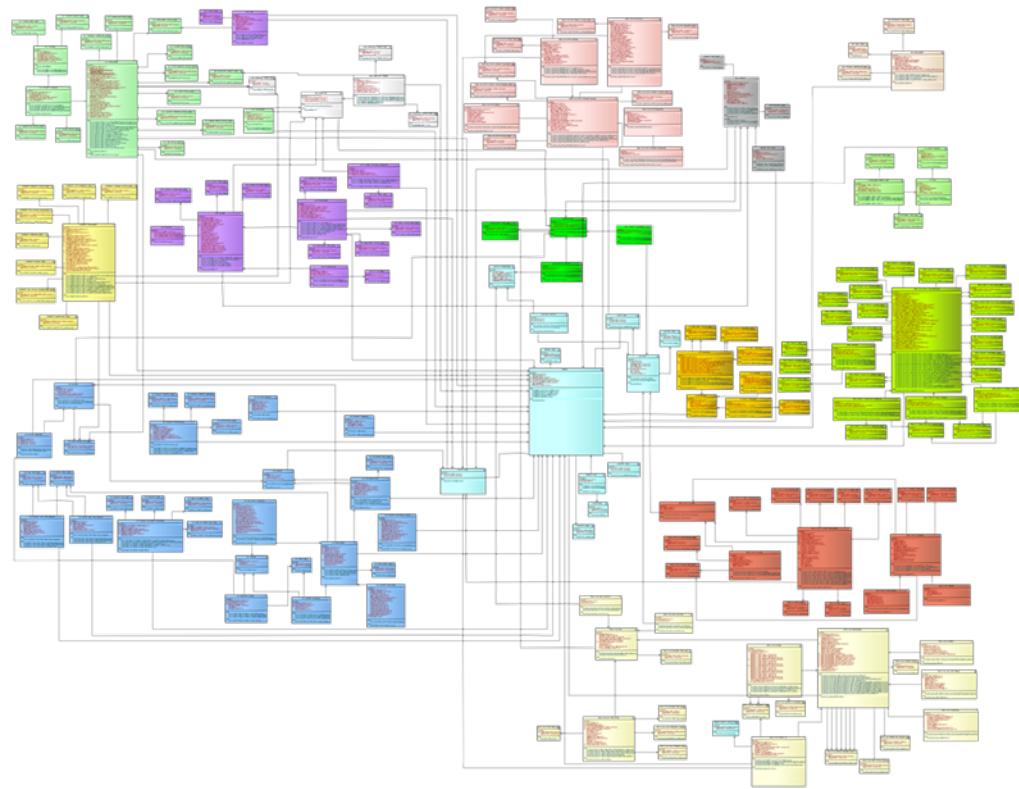
PROGRAMS:

- K12 Enrollment Program
- Workforce Course/Training
- WIA Enrollment



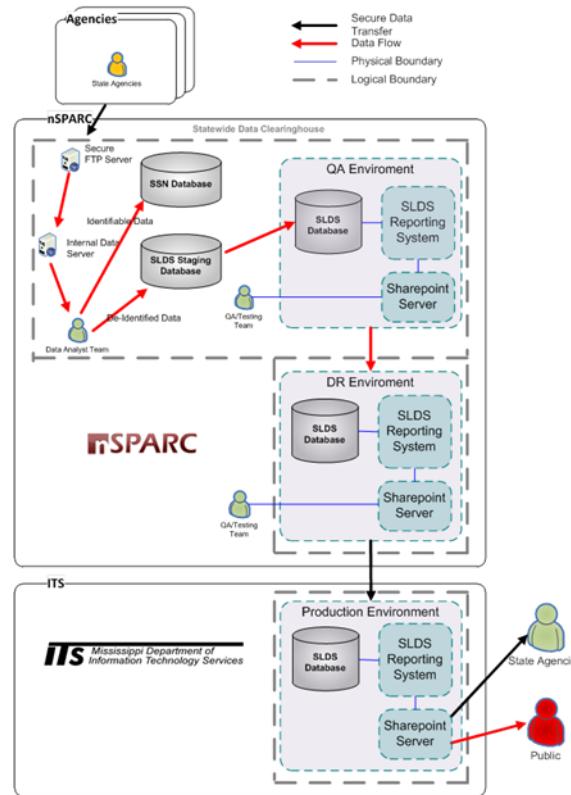
MISSISSIPPI

SLDS ENTITY RELATIONSHIP DIAGRAM



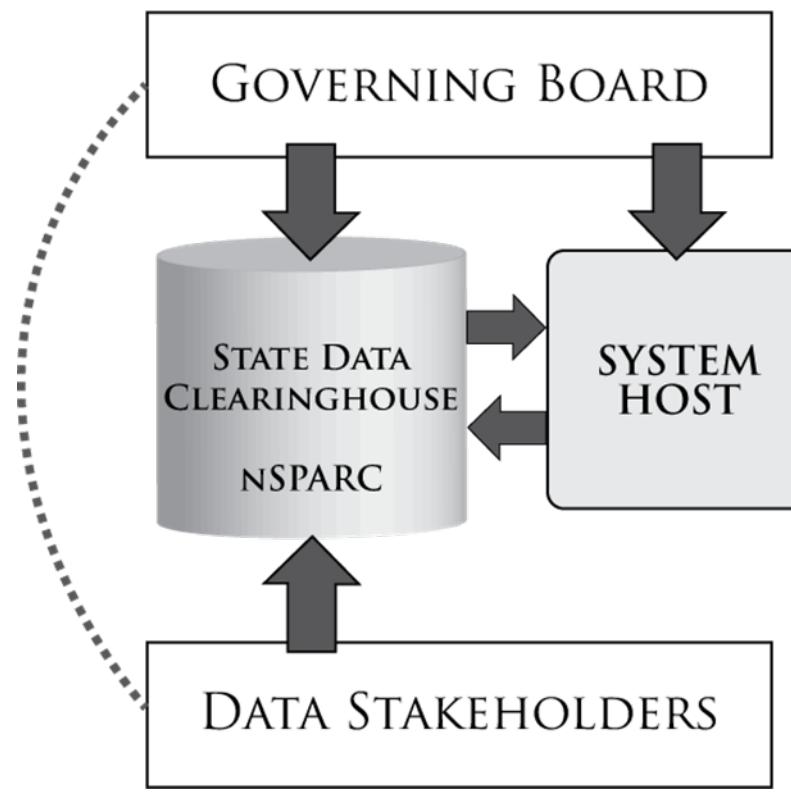
MISSISSIPPI

SLDS DATA FLOW ARCHITECTURE



MISSISSIPPI

GOVERNANCE STRUCTURE



MISSISSIPPI

DATA LIFECYCLE

AUDITING & MONITORING

DATA
TRANSFERRING

DATA
INVENTORY

DATA
VALIDATION

DATA
DE-IDENTIFICATION

DATA
ACCESS

RISK MANAGEMENT & TRAINING



DATA ACCESS



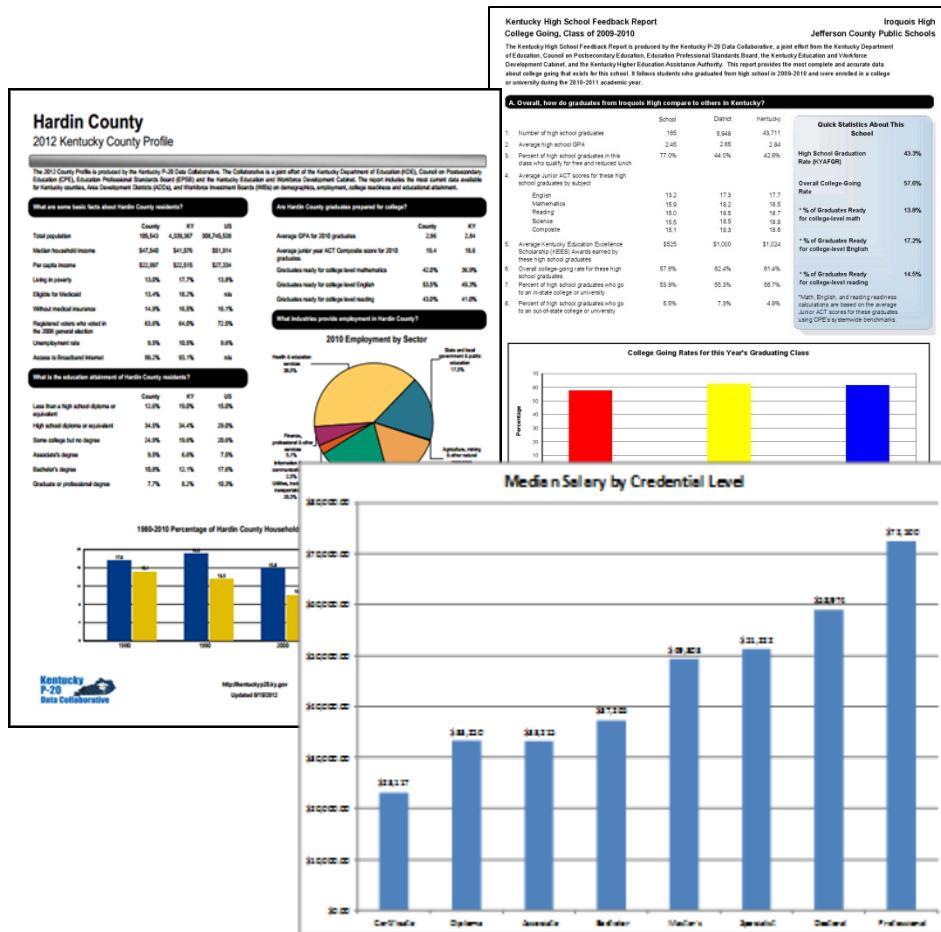
KENTUCKY

- Agencies have access to all the identifiable data they have provided to P-20 – but not to each other's.
- Agencies and P-20 Staff have access to the de-identified production data. Shared “universes” go live in December.
- P-20 Staff respond to multi-agency data requests with vetting process to validate for accuracy.
- Access through Business Objects Web Intelligence (WeBI), P-20 staff use WeBI, Crystal, SSRS, SPSS, SAS, Arc-View, and other tools.
- Data retention as needed. Long-term retention to be determined.



KENTUCKY

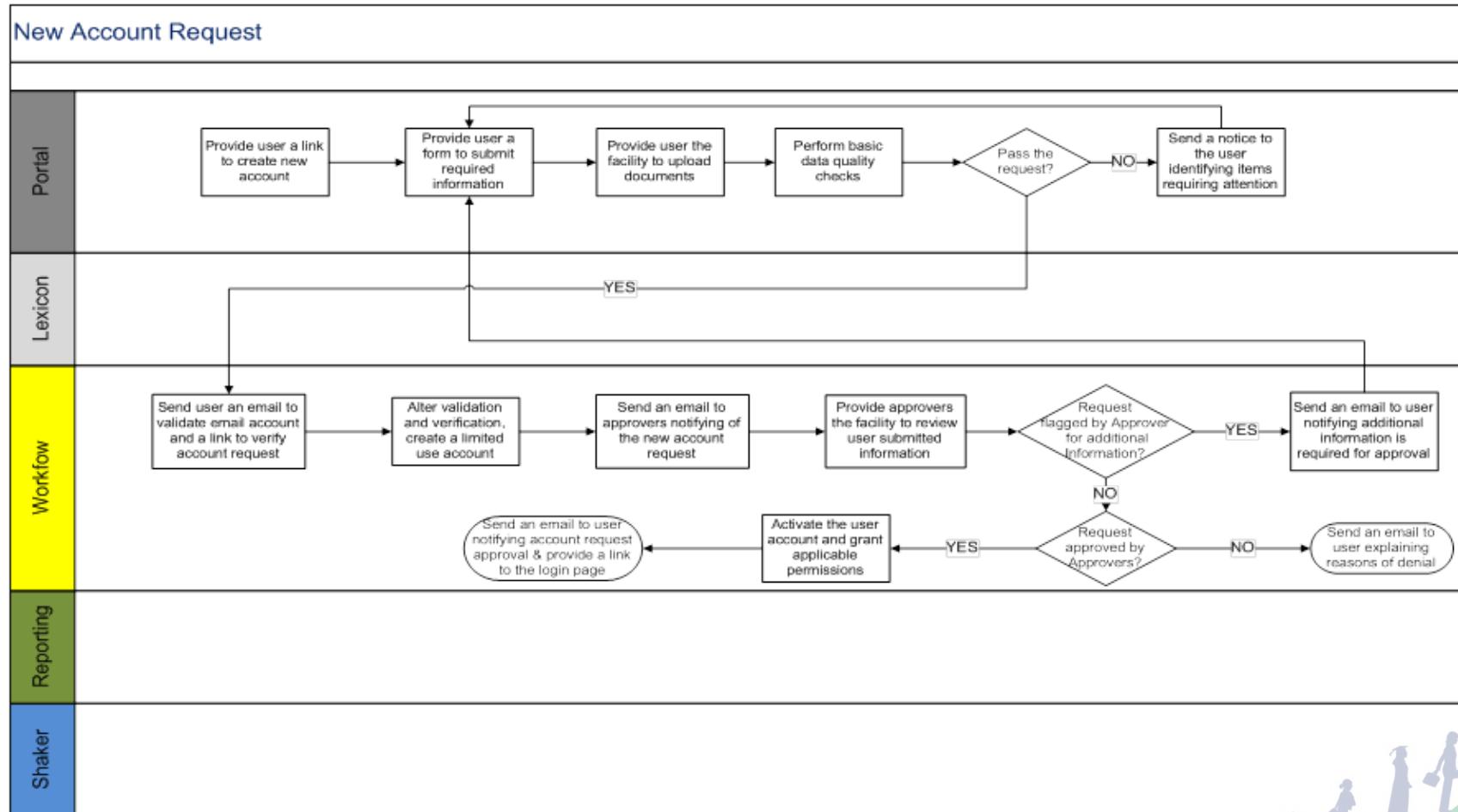
Centralized Reporting for Cross-Agency Issues



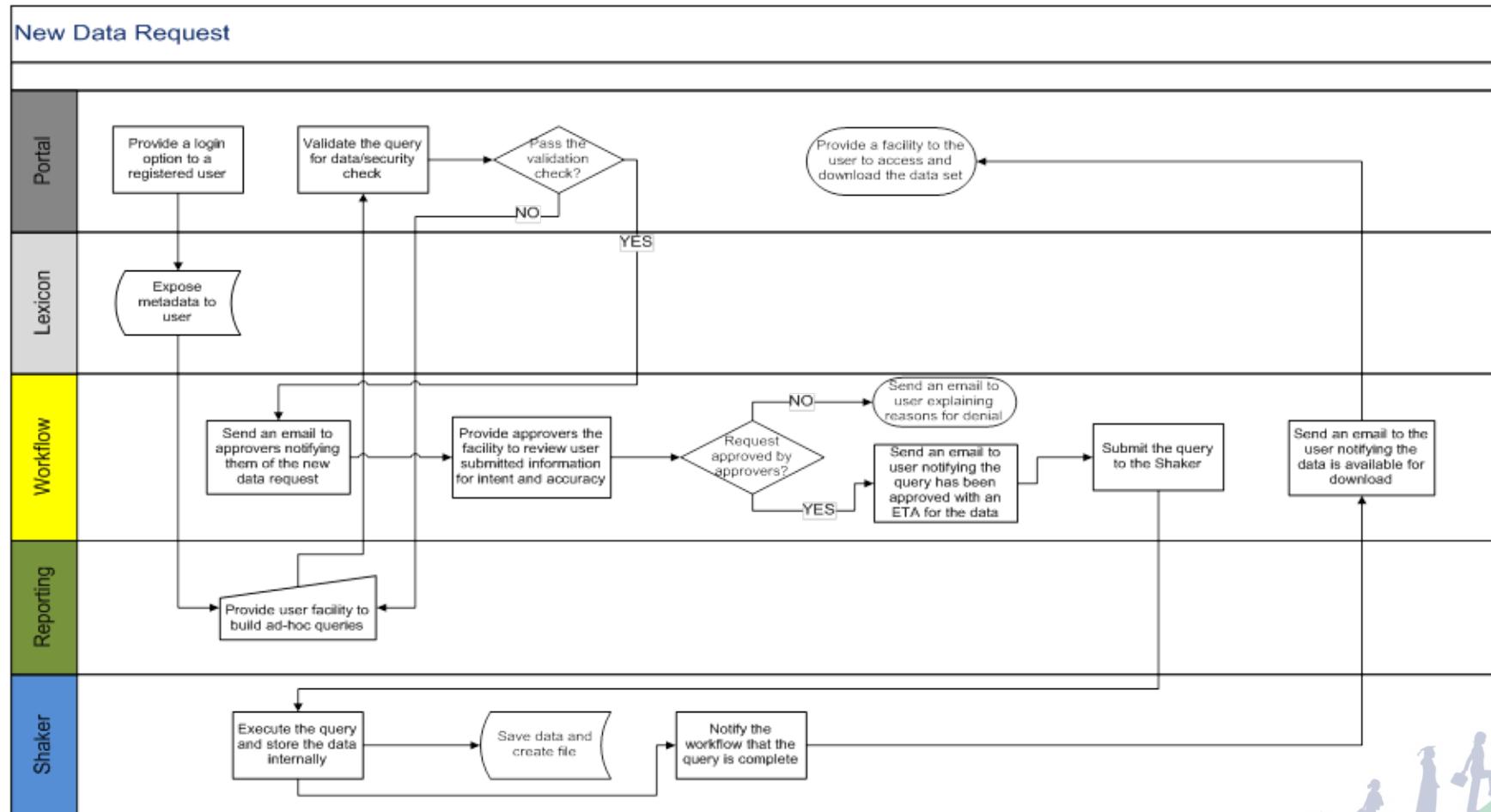
- High School Feedback
- Adult Education Feedback
- Employment Outcomes and Earnings
- County Profiles
- Workforce and Training Outcomes



VIRGINIA



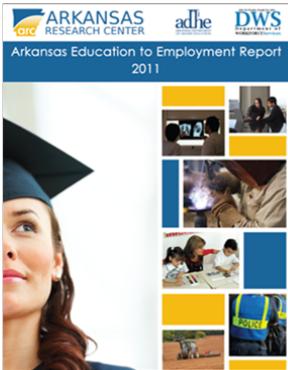
VIRGINIA



ARKANSAS

TrustEd: KIM & TIM

Employment Report Economic Success Measures



AIMS Report

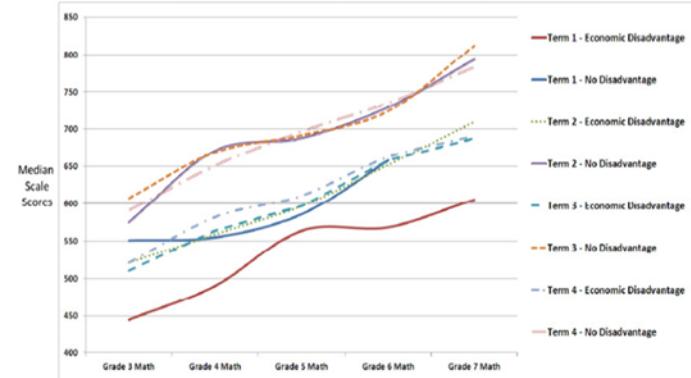
ACT	Concurrent Credit	AP Participant	Overall	On Time Bachelor's	On Time Bach. %
17-18			2438	586	24.0%
17-18		Y	553	189	34.2%
17-18	Y		222	73	32.9%
17-18	Y	Y	54	24	44.4%
19-20			2688	855	31.8%
19-20		Y	1024	434	42.4%
19-20	Y		457	184	40.3%
19-20	Y	Y	158	82	51.9%
21-22			2185	863	39.5%
21-22		Y	1376	683	49.6%
21-22	Y		483	238	49.3%
21-22	Y	Y	228	141	61.8%
23-24			1544	665	43.1%
23-24		Y	1615	931	57.6%
23-24	Y		465	261	56.1%
23-24	Y	Y	292	182	62.3%
25-26			938	471	50.2%
25-26		Y	1618	1013	62.6%
25-26	Y		371	224	60.4%
25-26	Y	Y	338	230	68.0%
27-29			662	365	55.1%
27-29		Y	1843	1285	69.7%
27-29	Y		306	196	64.1%
27-29	Y	Y	377	283	75.1%

Early Learning Report

KG Year	PreK_Summary Students	National Percentile Ranking (averaged)					
		G0_Lit	G0_Math	G0_Reading	G1_Lit	G1_Math	G1_Reading
2005	ABC	47	63.55	63.77	56.64	66.93	64.78
2005	HeadStart	152	57.28	60.67	51.70	61.09	59.93
2006	ABC	131	65.36	69.70	56.18	62.77	64.77
2006	HeadStart	126	61.34	60.45	55.03	54.67	60.22
2006	Voucher	18	51.50	53.94	46.39	46.88	49.19
2007	ABC	291	59.93	64.96	55.07	47.29	51.39
2007	HeadStart	108	58.41	64.77	51.47	44.73	49.57
2007	Voucher	15	47.53	43.53	44.27	48.67	50.67
2008	ABC	325	51.61	58.36	45.35	52.57	
2008	HeadStart	107	50.23	57.52	43.47	47.33	
2008	Voucher	25	45.96	57.40	44.96	40.32	
2009	ABC	540	51.65	54.76	48.68	55.97	
2009	HeadStart	129	51.71	54.78	47.95	51.72	
2009	Voucher	26	47.04	49.19	37.26	46.87	
2010	ABC	720	53.18	55.95	57.70	56.59	
2010	HeadStart	120	49.29	52.16	56.12	54.99	
2010	Voucher	33	46.12	54.55	48.48	55.87	
2011	ABC	750	81.64	67.73	76.05		
2011	HeadStart	145	78.10	62.64	69.92		
2011	Voucher	44	78.82	62.16	73.30		

UAMS Neonatal Report

1998 UAMS Births - ADE Assessments 2007-2011



DATA ACCESS

- ON REQUEST

- ONE-STOP PORTA



MISSISSIPPI

IMPACT OF READING PROFICIENCY

- For students not proficient in reading in 8th grade, what is their likelihood to succeed as young adults?
 - 70 percent less likely to graduate from high school
 - 65 percent less likely to go to college
 - 3.7 times more likely to take remedial courses
 - 2.7 times more likely to receive food stamps or TANF
 - 3.7 times more likely to go to prison
- Total Annual Average Cost to Mississippi: \$143 Million

ON REQUEST

▼ EXAMPLES

• IMPACT OF READING PROFICIENCY

• JOB CREATION

• WORKFORCE NEEDS

• ALIGNING COLLEGE GRADS WITH WORKFORCE

• ALIGNMENT OF EDUCATION SYSTEMS

• TRACKING STUDENT OUTCOMES

• LINKING TEACHERS TO STUDENT OUTCOMES

• LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI

JOB CREATION

WORKFORCE PIPELINE

- Sector: Transportation and Logistics
- Potential Site: Desoto County
- Workforce pipeline: students enrolled in the postsecondary system graduating within a year.

	COMMUNITY COLLEGE	PUBLIC UNIVERSITY
Computer and Information Sciences and Support Services.	472	137
Transportation and Materials Moving.	19	0
Business, Management, Marketing, and Related Support Services	1,278	1,010
TOTAL...	1,769	1,147

Source: Mississippi State Longitudinal Data System, 2012

ON REQUEST

EXAMPLES

- IMPACT OF READING PROFICIENCY

- JOB CREATION

- WORKFORCE NEEDS

- ALIGNING COLLEGE GRADS WITH WORKFORCE

- ALIGNMENT OF EDUCATION SYSTEMS

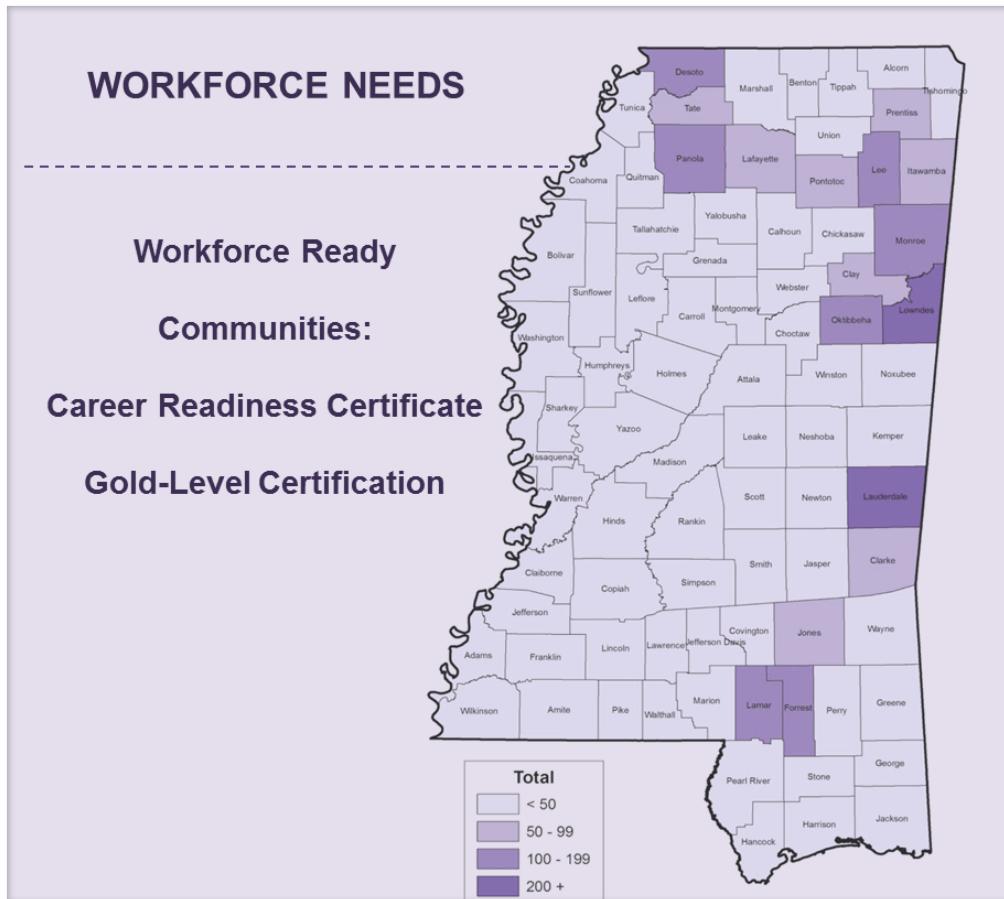
- TRACKING STUDENT OUTCOMES

- LINKING TEACHERS TO STUDENT OUTCOMES

- LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI



- ON REQUEST**
- ▼ EXAMPLES
- IMPACT OF READING PROFICIENCY
 - JOB CREATION
 - WORKFORCE NEEDS
 - ALIGNING COLLEGE GRADS WITH WORKFORCE
 - ALIGNMENT OF EDUCATION SYSTEMS
 - TRACKING STUDENT OUTCOMES
 - LINKING TEACHERS TO STUDENT OUTCOMES
 - LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI

ALIGNING COLLEGE GRADUATES WITH THE WORKFORCE

- 74 percent of university graduates stay and work in Mississippi
- 79 percent of community college graduates stay and work in Mississippi
- University Majors most likely to stay:
 - Education
 - Healthcare
 - Public Administration
- University Majors Most likely to leave
 - Physical Sciences
 - Math and Statistics
 - Engineering

ON REQUEST

▼ EXAMPLES

- IMPACT OF READING PROFICIENCY
- JOB CREATION
- WORKFORCE NEEDS
- ALIGNING COLLEGE GRADS WITH WORKFORCE
- ALIGNMENT OF EDUCATION SYSTEMS
- TRACKING STUDENT OUTCOMES
- LINKING TEACHERS TO STUDENT OUTCOMES
- LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI

ALIGNMENT OF EDUCATION SYSTEMS

- 69 percent of all high school graduates go to college
 - 54 percent go to a community college
 - 17 percent go to a university
- 50 percent enroll in remedial math and English courses
 - 60 percent of community college students take remedial courses
 - 28 percent of university students take remedial courses

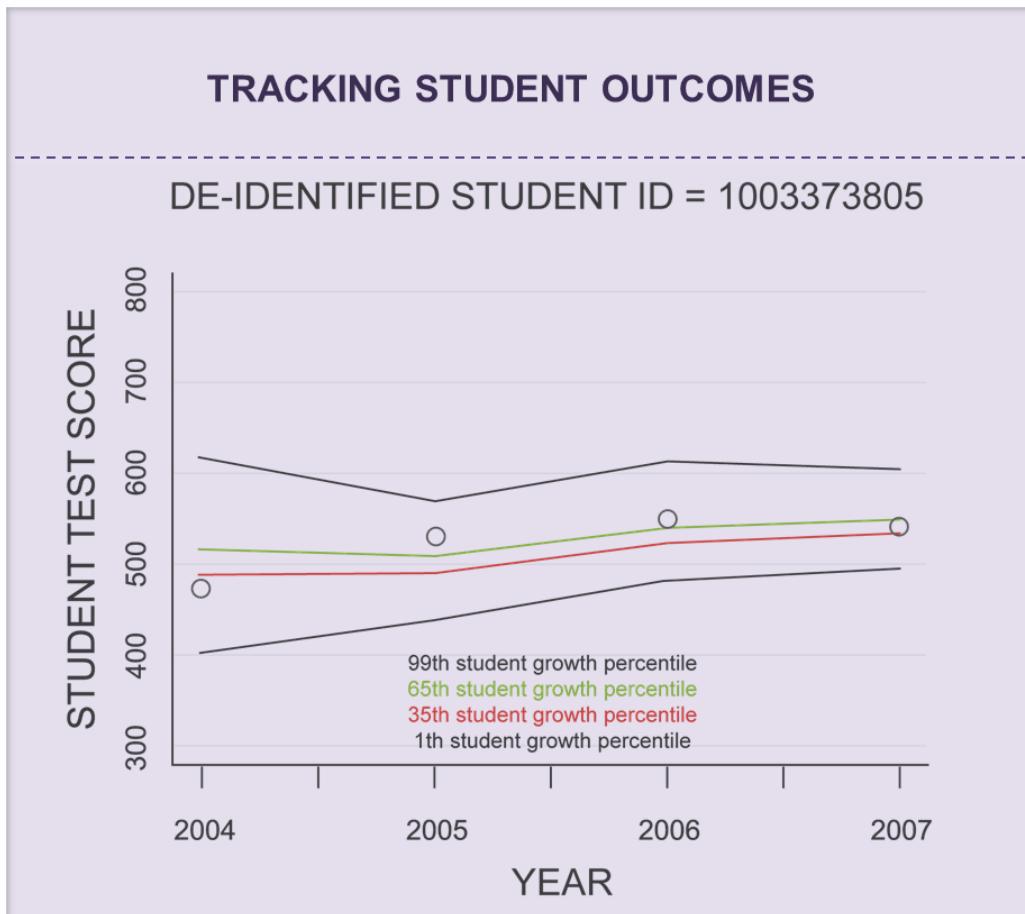
ON REQUEST

▼ EXAMPLES

- IMPACT OF READING PROFICIENCY
- JOB CREATION
- WORKFORCE NEEDS
- ALIGNING COLLEGE GRADS WITH WORKFORCE
- ALIGNMENT OF EDUCATION SYSTEMS
- TRACKING STUDENT OUTCOMES
- LINKING TEACHERS TO STUDENT OUTCOMES
- LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI



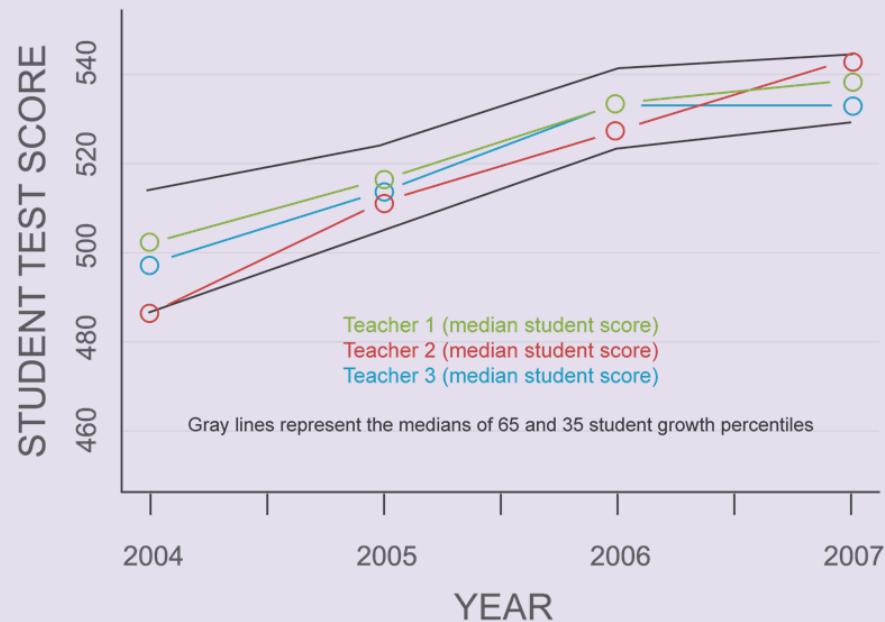
- ON REQUEST
- ▼ EXAMPLES
- IMPACT OF READING PROFICIENCY
 - JOB CREATION
 - WORKFORCE NEEDS
 - ALIGNING COLLEGE GRADS WITH WORKFORCE
 - ALIGNMENT OF EDUCATION SYSTEMS
 - TRACKING STUDENT OUTCOMES
 - LINKING TEACHERS TO STUDENT OUTCOMES
 - LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI

LINKING TEACHERS TO STUDENT OUTCOMES

COMPARISON OF 3 TEACHERS



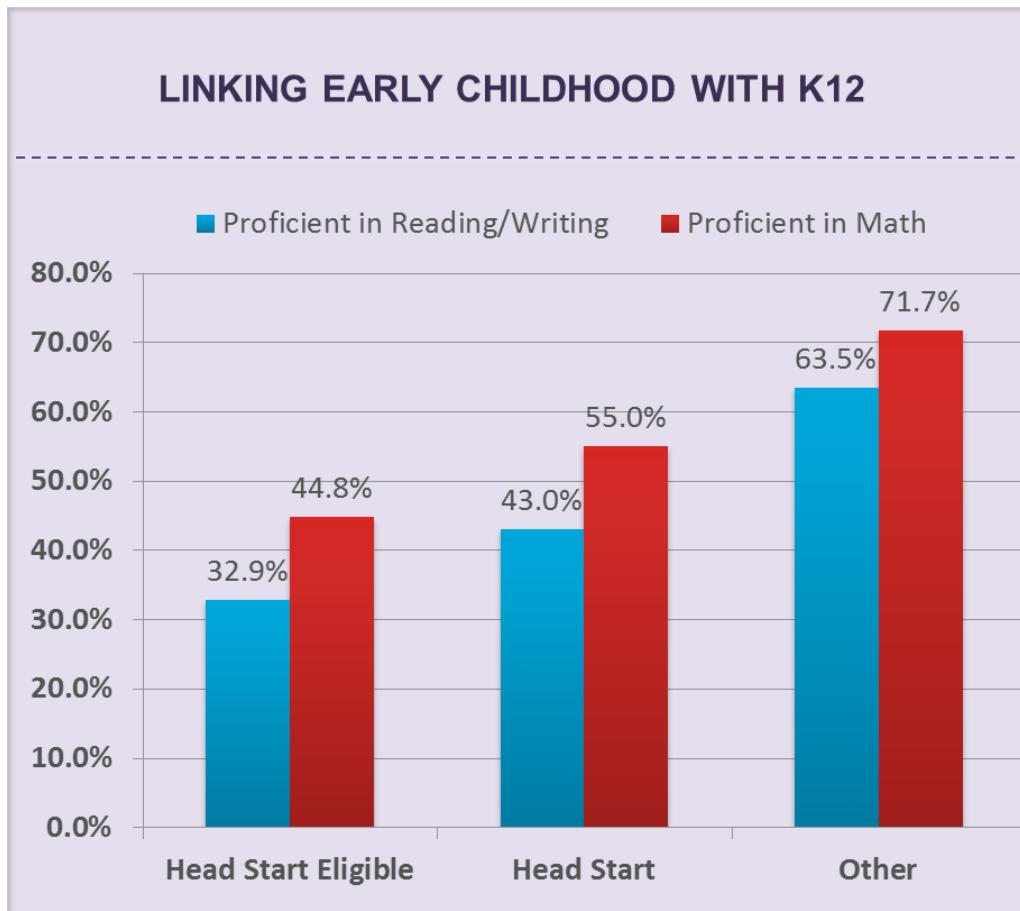
ON REQUEST

EXAMPLES

- IMPACT OF READING PROFICIENCY
- JOB CREATION
- WORKFORCE NEEDS
- ALIGNING COLLEGE GRADS WITH WORKFORCE
- ALIGNMENT OF EDUCATION SYSTEMS
- TRACKING STUDENT OUTCOMES
- LINKING TEACHERS TO STUDENT OUTCOMES
- LINKING EARLY CHILDHOOD WITH K12



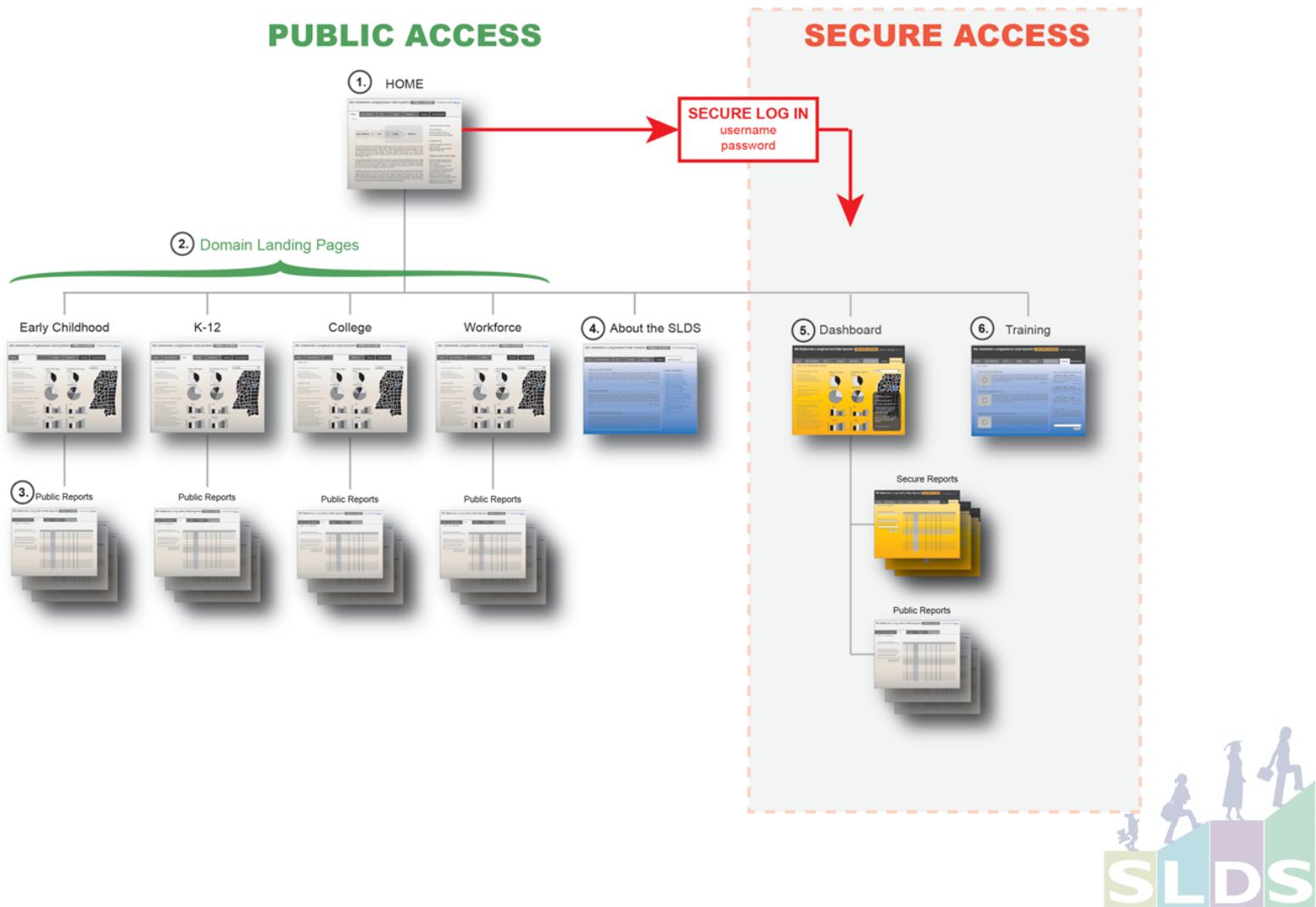
MISSISSIPPI



- ON REQUEST
- ▼ EXAMPLES
- IMPACT OF READING PROFICIENCY
 - JOB CREATION
 - WORKFORCE NEEDS
 - ALIGNING COLLEGE GRADS WITH WORKFORCE
 - ALIGNMENT OF EDUCATION SYSTEMS
 - TRACKING STUDENT OUTCOMES
 - LINKING TEACHERS TO STUDENT OUTCOMES
 - LINKING EARLY CHILDHOOD WITH K12



MISSISSIPPI



CONTACTS & ADDITIONAL RESOURCES

Contact information:

Charles McGrew, Charles.McGrew@ky.gov

Aaron Schroeder, aaron.schroeder@vt.edu

Neal Gibson, Neal.Gibson@arkansas.gov

Mimmo Parisi, MParisi@nsparc.msstate.edu

Jim Campbell, jim.campbell@sst-slds.org

Jeff Sellers, jeff.sellers@sst-slds.org

Keith Brown, keith.brown@sst-slds.org

Resources:

http://nces.ed.gov/programs/slds/pdf/federated_centered_print.pdf

