

Data Management for Large STEP Projects

Michigan State University & Lansing Community College
NSF STEP PI Meeting March 15, 2012
Workshop Session I-08

Mark Urban-Lurain, MSU

Daina Briedis, MSU

Renée DeGraaf, LCC

Ruth Heckman, LCC

Colleen McDonough, MSU

Jon Sticklen, MSU

Claudia Vergara, MSU

Tom Wolff, MSU

Acknowledgements

This material is based upon work supported by the National Science Foundation under award 0757020 (DUE). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF).



Workshop Structure

- Goal of the workshop is for participants to help identify some common challenges and successful solutions that can be shared across the STEP community
- Working in groups
- Identify characteristics of your project
- Reporting results for synthesis
- Summary provided to all participants after workshop

Who are we?

- EEES: Engaging Early Engineering Students
 - Targeted disciplines
 - All departments in College of Engineering
 - Targeted numbers of students
 - All incoming engineering students (~1000 first year students/year)
 - How long project in place
 - 4 years
 - Types of interventions
 - Peer Assisted Learning
 - Connector Faculty
 - Dx-Driven Early Intervention
 - Cross Course Linkages

Who are you? At your tables...

- Your name and institution
- Quickly tell everyone about your project
 - Targeted discipline(s)
 - Targeted numbers of students
 - What year of your STEP project are you in?
 - Types of intervention(s)
- Why are you attending the workshop?
 - Each table will report one reason for attending the workshop

Why are You Attending This Workshop?

Part 1: Data Collection

- Why are we collecting data?
- Research/evaluation challenges?
- What types of analyses?

Part 1 EEES:

Why are we collecting data?

- STEP Project Evaluation
 - Which students / faculty participate in which interventions?
 - What are the barriers/inducements to participation?
 - How can each intervention be improved?
- Research
 - What are the changes in retention as a result of EEES?
 - What are the impacts of each of the components of EEES on retention?
 - Which groups of students are impacted by which components of EEES?
 - What are the interactions among the EEES components for different groups of students?

Part 1 EEES: Analyzing Data

- Research/evaluation challenges
 - Interactions
 - Different types of at-risk students
 - Different interventions
 - Final outcomes
 - College-wide
 - Not feasible to compare contemporaneous students who are and are not participating in the interventions
 - Elective participation in parts of the program (or not)
 - Interactions may vary by student
 - Traditional statistical techniques difficult to use
- Types of analyses we are performing
 - Structural Equation Modeling
- Software we are using
 - AMOS

Part 1: Data Collection

- Why are you collecting data?
 - STEP Project Evaluation
 - What are your evaluation questions/metrics?
 - Research
 - What are your research questions?
- Analyzing Data
 - What are your research/evaluation challenges?
 - What types of analyses are you performing?
 - What software are you using?
- Identify most common themes to report

Part 1: Report Out

Part 2: Types of Data

- What are the sources?
 - What data is from students?
 - What data is from faculty
 - What data is about the course/instruction/intervention overall?
 - Other data?
- Type of data
 - Quantitative
 - Qualitative
- Form of the data:
 - Electronic
 - Paper
- Who enters / transcribes the data?
 - Participants directly (e.g., online)
 - Researchers

Part 2: EEES Data

- Student data
 - Demographic data (class standing, gender, ethnicity, ACT scores)
 - From Registrar's office
 - Assignment, exam and course grades in the core pre-engineering courses
 - From faculty
 - Grades in other courses
 - From Registrar's office
 - Math diagnostic exam scores
 - From Math department
 - Number of PAL sessions attended
 - Student sign-in sheets
 - Number of meetings with Connector Faculty
 - Surveys of students and faculty
 - Student perceptions of the various components of the EEES program
 - Surveys and interviews
 - Student engagement and desire to persist in engineering
 - Surveys and interviews
- Outcome data (From Registrar's office and College)
 - Students who graduate with an engineering degree
 - Students who are academically qualified for engineering but elect to leave for other majors
 - Students who apply but are not admitted due to insufficient grades

Part 2: Metadata

- What descriptors are you collecting for your data?
- Any “standards” (e.g., ABET, etc.)

Part 2: Identifiable Data

- Why do you want identifiers?
 - Allows merging/tracking for given individuals across data sets
 - Pre-test / post-test
 - Across courses
 - Analysis across different types of data
- Which of your data has identifiers?
- Anonymizing data
 - Protects identity
 - Needs to be consistent
 - Same identifier gets same code across data

Part 2: Your Data

- Identify the data you are collecting
- What are the sources?
 - What data is from students?
 - What data is from faculty
 - What data is about the course/instruction/intervention overall?
 - Other data?
- Type of data
 - Quantitative
 - Qualitative
- Form of the data:
 - Electronic
 - Paper
- Who enters / transcribes the data?
 - Participants directly (e.g., online)
 - Researchers
- What Metadata
- Identifiers
 - What are they
 - Anonymizing/protecting

Part 2: Report Out

Part 3: Managing EEES Data

- Strategies for managing data
 - One researcher coordinates data collection
 - Master list of data to be collected, source, timetable, responsible person
- What software are you using?
 - Quantitative data in spreadsheets
 - Paper scanned to PDF
 - Audio recordings of interviews as MP3
 - Merging data sets in SPSS Modeler
- Where are you storing the data (local computers, cloud, etc.)
 - Networked file storage on MSU campus
- How do you control access?
 - Restricted access to data manager
 - Parceled out to individual research assistants for cleanup
- Metadata
 - Data dictionary describing each set of data
 - File naming conventions
 - Directory structures for incoming “raw” data and cleaned up data ready for analysis
- Anonymization of identified data
 - SHA-1 Hashing of student identifiers

Part 3: Managing Your Data

- What are your strategies for managing data?
- What software are you using?
- Where are you storing the data (local computers, cloud, etc.)
- How to you control access?
- Metadata
- Anonymization of identified data

Part 3: Report Out

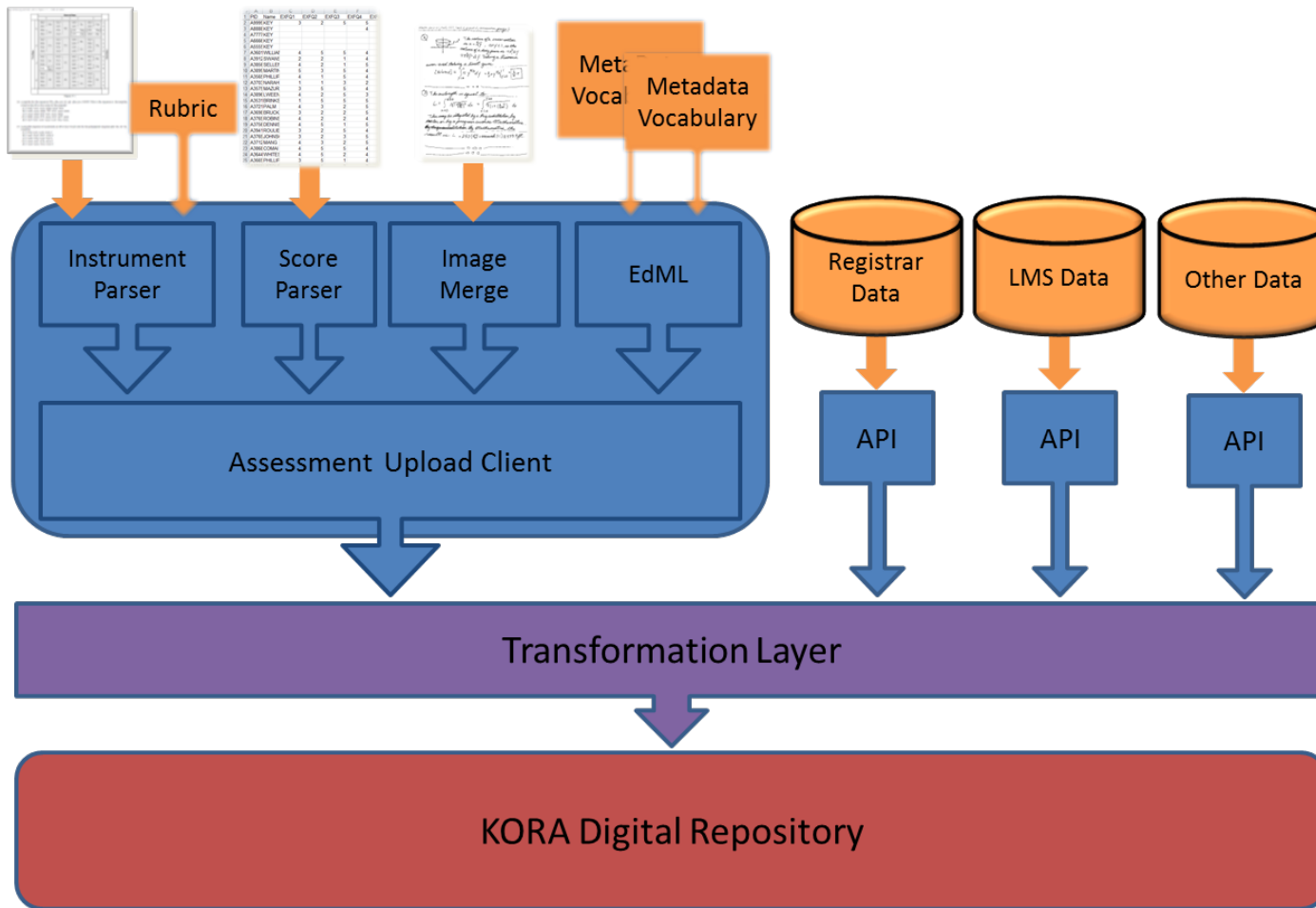
Challenge: Managing Diverse Datasets

- Need to support any type of assessments
- Independent of any LMS
- Data granularity to item level
- Variety of meta-data standards
- Merging/tracking data across courses/programs within the institution
- Supporting research across institutions

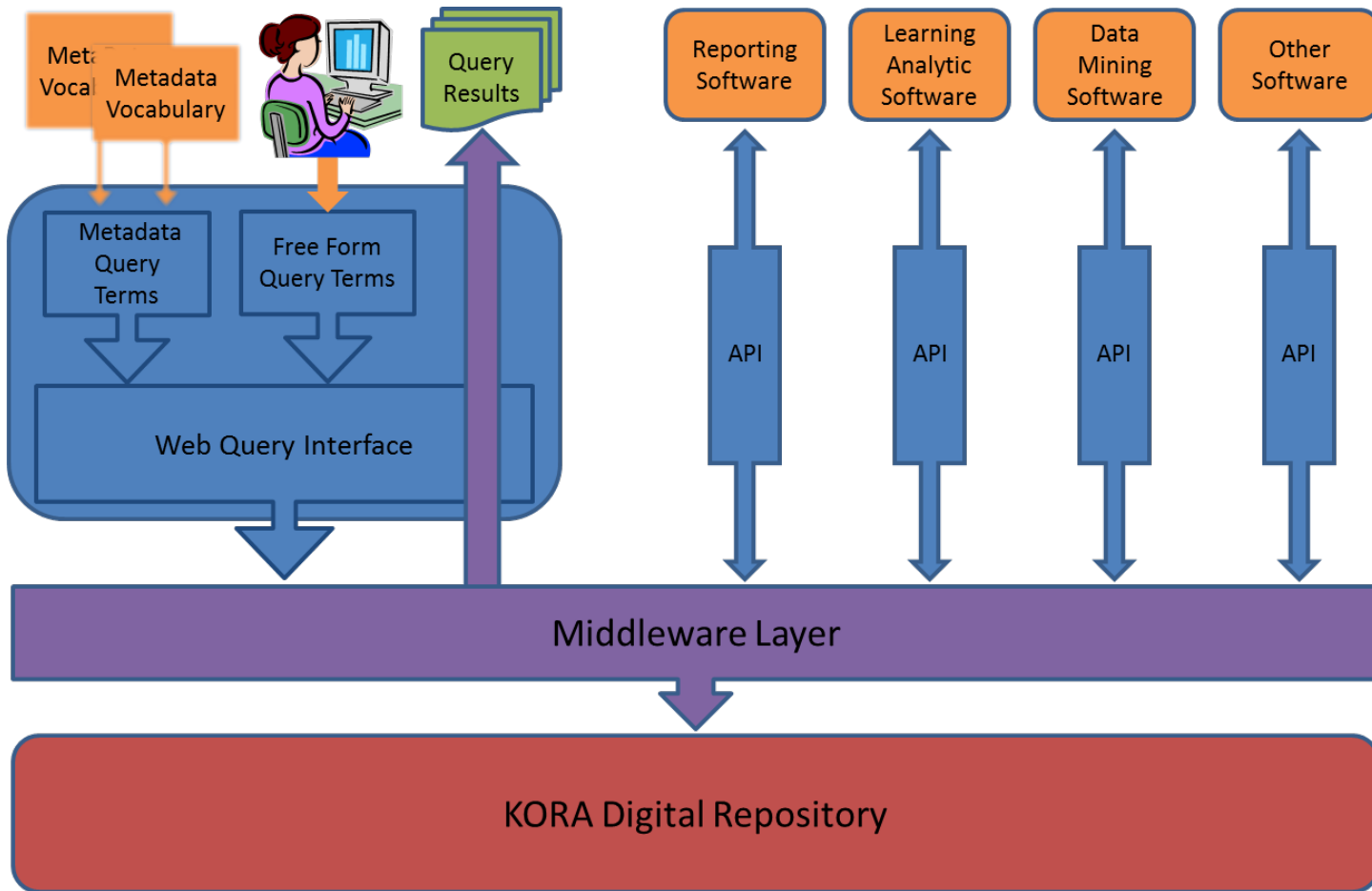
Learning Outcomes Assessment Data (LOAD) Store Proposal

- Developed prototype data store
 - NSF 0618501
- Store student assessment data at the item level
 - Individual student responses on each assessment item
- Merge data from multiple sources
 - Across courses
 - Institutional data
- Educational Metadata Standards (EdML)
 - Built on existing standards (Dublin Core, IMS Question Test Interoperability specification)
- Exam parsing
 - Extract individual exam items from PDF and store

LOAD Data Upload



LOAD Data Query



Thanks for Your Participation

- Questions?
 - Mark Urban-Lurain urban@msu.edu
 - Claudia Vergara vergara@msu.edu