

MINING BIG DATA TO SOLVE THE RETENTION AND GRADUATION PUZZLE

Patrice Lancey, Ph.D.

Uday Nair, M.S, MBA

Rachel Straney, M.S.

Association for Institutional Research
2014 Forum
May 30, 2014



University of Central Florida

OVERVIEW

- UCF characteristics and graduation rate goal
- OEAS existing capabilities and challenges
- building a student success database (longitudinal)
- examples of data mining findings
- implications and future work

UNIVERSITY OF CENTRAL FLORIDA

- large public metropolitan research university with almost 60,000 enrolled students
- Carnegie Classification – Research University with Very High Research Activity (RU/VH)
- 12 colleges including a medical school
- 211 degree programs (91 bachelor's, 86 master's, 3 specialist, 30 doctoral, 1 professional)
- approximately 45% of UCF students are transfer students

THE GOAL

Institution

- increase six year graduation rates from 63% (2011-12) to 70% in five years

Assessment and Analytical Study Support Office

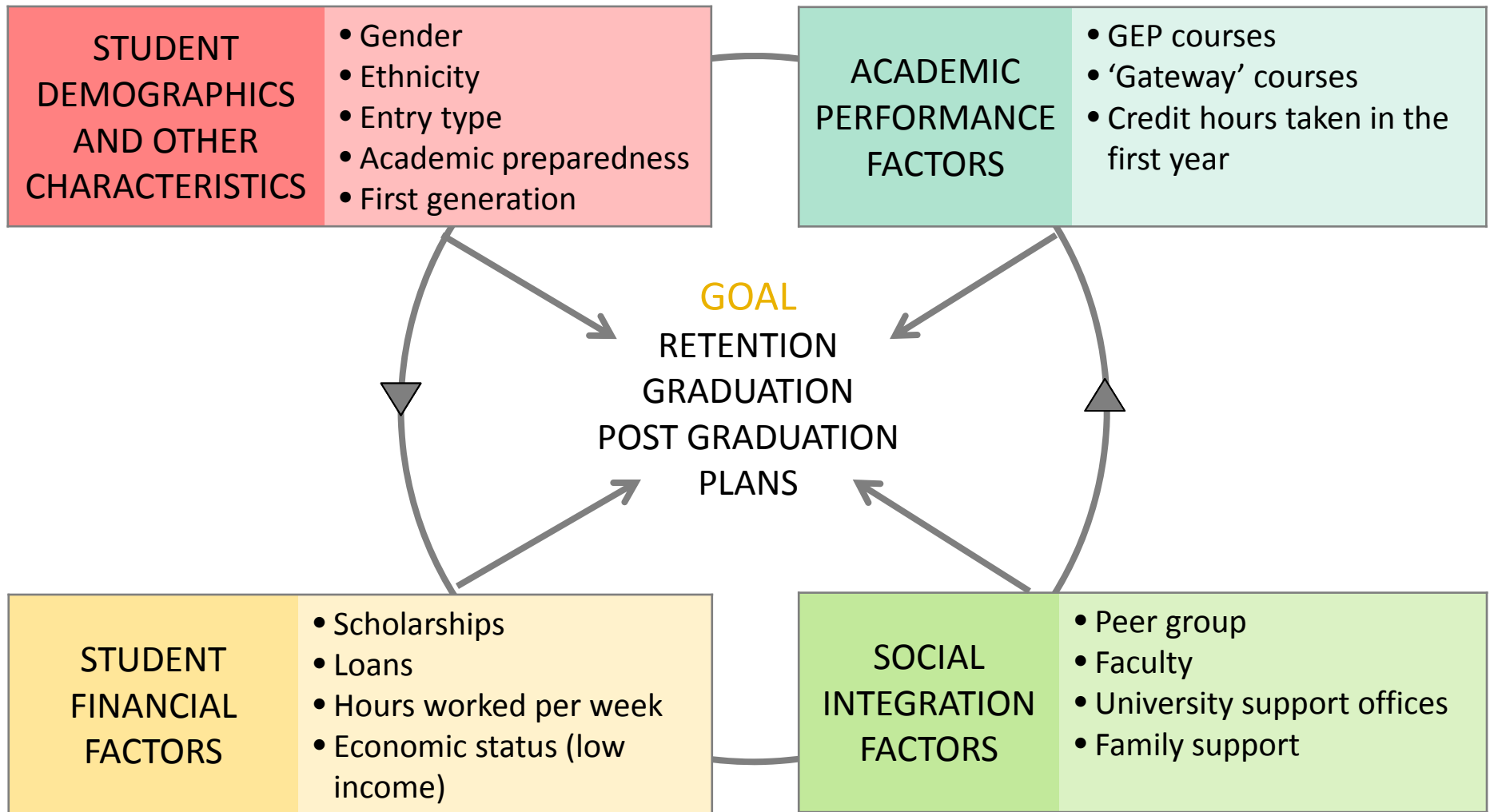
- build on existing capabilities to design and carry out analytical studies to support university strategic directions
 - design and develop a comprehensive longitudinal data base
 - mine data for findings to develop strategies that will increase retention and graduation rates for prevalent profiles of students

STUDENT SUCCESS MODEL*

- factors considered for the student success model are supported in the literature (Tinto, Astin & Scherrei, Berger & Braxton)
- recent conversations at many levels have focused on goal attainment – as defined by improved graduation rates and successfully attaining post graduation goals
- the student success model is dynamic, complex and should include interactive factors

* Adapted from Tinto's model and other work related to student success.

STUDENT SUCCESS MODEL*



* Adapted from Tinto's model and other work related to student success.

EXISTING CAPABILITIES: INTEGRATION OF SURVEY AND INSTITUTIONAL DATA SOURCES

- tradition of producing reports that merge student survey results with student institutional data elements
- mature survey database preparation, processing and reporting business procedures
- experience building custom cohort data sets for specific analytical research projects to model student success
- practice using institutional data tables and elements stored in PeopleSoft and the UCF SAS warehouse

EXISTING CHALLENGES

- custom database build for each research project
- a standardized process did not exist
- inefficient use of staff time
- lag in response time

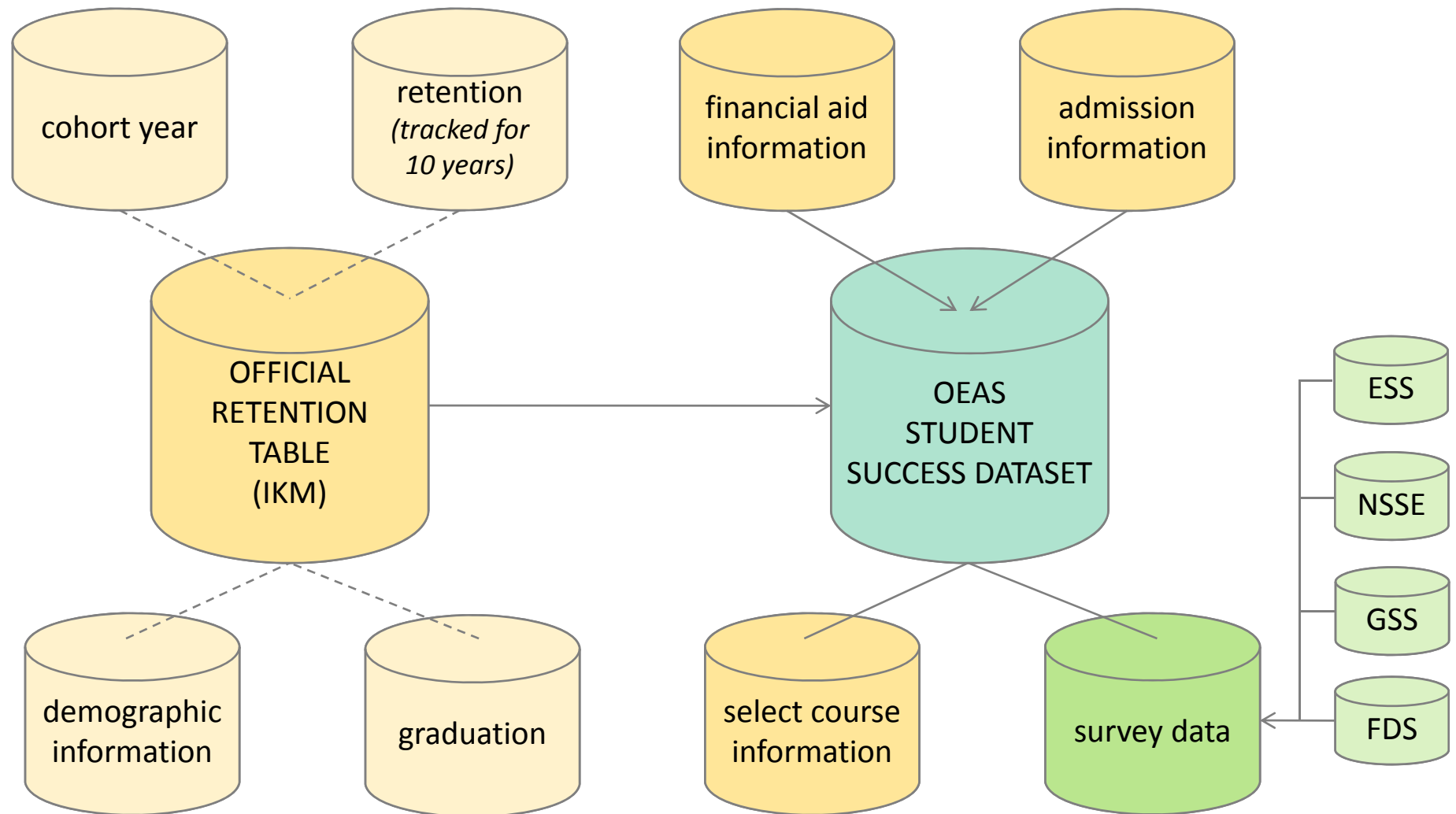
REQUIREMENTS FOR THE COMPREHENSIVE LONGITUDINAL DATABASE

- all students entering the university should be tracked irrespective of their graduation status
- comprehensive data set should include variables that could potentially measure factors associated with student outcomes
- final data set should have a dimension of student-level to make it easy to analyze and interpret
- data set should be easily processed on continual basis

CHALLENGES TO CREATING A COMPREHENSIVE LONGITUDINAL DATABASE

- data dimension for official student records vary:
 - student level: demographic, characteristics, etc.
 - student-term level: financial aid, course load taken, etc.
 - student-term-course level: course grades
- some of the data are available in official records while others can only be obtained through other sources such as self-reporting, student tracker, etc.
- data are available at different points in time during a student's enrollment at UCF
- data are in various "siloes"

CONCEPTUAL FRAMEWORK OF THE COMPREHENSIVE LONGITUDINAL DATABASE



EXISTING CAPABILITIES: INSTITUTIONAL SURVEY SOURCES

- Entering Student Survey (ESS)
- National Survey of Student Engagement (NSSE)
- Graduating Student Surveys (GSS)
- First Destination Survey (FDS)

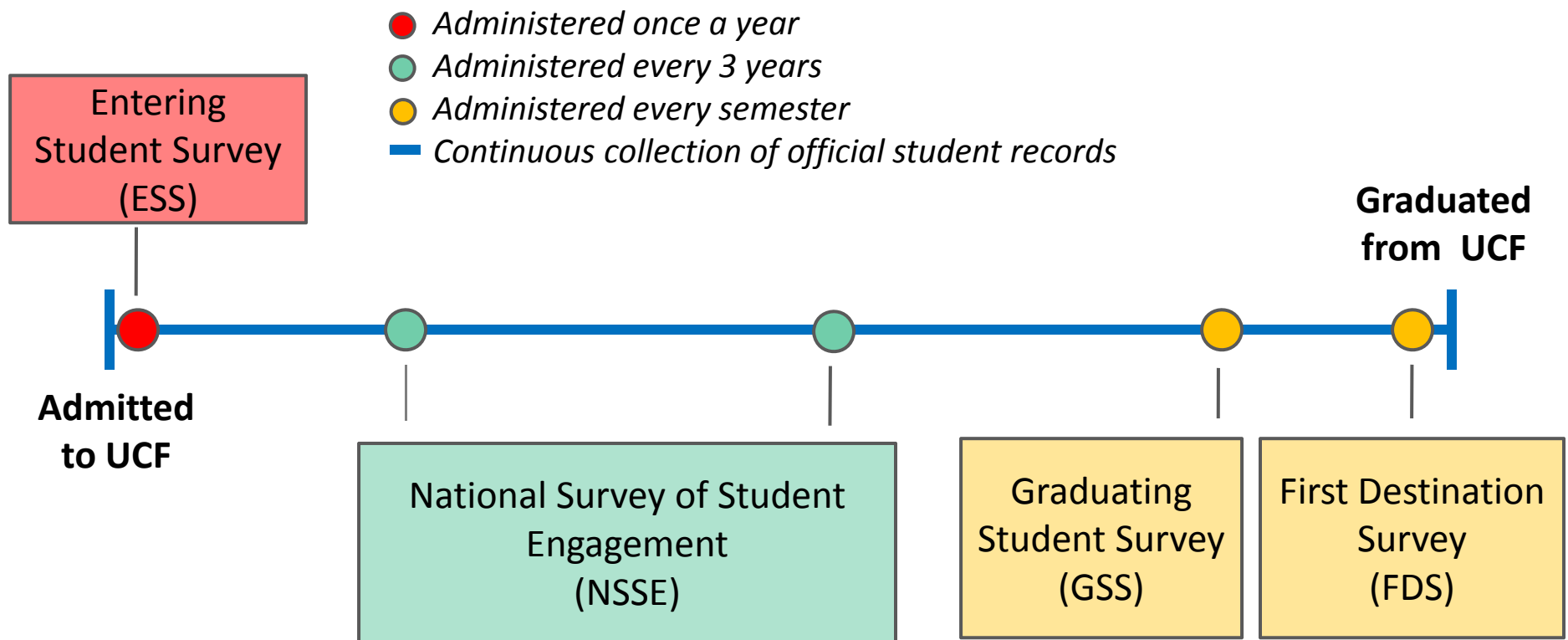
Key Features

disseminated through student portal or business process

high response rates

capture of student ID as primary key

STUDENT SURVEY TIMELINE AT UCF



STUDY 1: IDENTIFYING FACTORS ASSOCIATED WITH FTIC GRADUATION

- univariate logistic regression models were constructed using 2004-05 and 2005-06 Summer/Fall entering cohorts to predict 6-year graduation
- variables considered:
 - first-year retention
 - high school cumulative GPA
 - financial aid (loans, scholarships)
 - demographics (gender, ethnicity)
 - general education course performance at UCF

STUDY 1: MODEL DIAGNOSTICS

Diagnostic ranking for significant models

Variables	Model Diagnostics				Diagnostics Ranking			
	Wald Chi Square <i>Larger the better</i>	p-value <i>< 0.05 means the variable is significant</i>	Deviance (-2 Log L) <i>Smaller the better</i>	c statistic <i>Larger the better</i>	Wald Ranking	Deviance Ranking	c statistic Ranking	Combined Ranking
1 year retention	1646.9	< 0.0001	11855	0.713	2	6	2	10
GPA of GEP courses taken in year 1	1834.5	< 0.0001	12594	0.756	1	9	1	11
Bright Futures Scholarship awarded in year 1	382.7	< 0.0001	12744	0.613	9	10	9	28
HS GPA	593.5	< 0.0001	14741	0.630	8	14	8	30
Advanced Placement Credit	186.0	< 0.0001	15289	0.562	10	15	10	35
Grant/scholarship awarded in year 1	3.9	0.0458	6538	0.508	16	2	18	36
Bright Futures Scholarship awarded	130.0	< 0.0001	15352	0.530	11	16	12	39
Gender	67.1	< 0.0001	15414	0.539	12	17	11	40
SAT Math Score	11.6	0.0007	14724	0.519	14	12	14	40
SAT Total Score	9.4	0.0022	14718	0.516	15	11	16	42
Ethnicity	21.4	< 0.0001	15460	0.518	13	18	15	46

STUDY 1: KEY FINDINGS

- 11 out of 19 variables considered were found to be associated with 6-year graduation
- first year retention and GPA of GEP courses taken in the first year were the best predictors based on the combined ranking of model diagnostics
- high school GPA, and demographic variables were found to be statistically significant but with lower predictive power

STUDY 2: IDENTIFYING FACTORS ASSOCIATED WITH GRADUATION FOR TRANSFER STUDENTS WITH AA/AS DEGREES

- multivariate logistic regression models were constructed using 2008-09 and 2009-10 Summer/Fall entering cohorts
 - model 1: graduation (yes/no)
 - model 2: time to graduation (within/after 3 years)
- variables considered:
 - first generation in college
 - financial need
 - hours worked per week*
 - affiliated college upon entry to UCF
 - prior institution
 - demographics (gender, ethnicity, etc.)

* Predictor used in model 2 since hours worked was collected on the Graduating Student Survey

STUDY 2: MODEL 1

- factors associated with graduation include:
 - first generation in college (*p-value* < 0.0001)
 - affiliated college upon entry to UCF (*p-value* < 0.0001)
 - age (*p-value* < 0.0001)
 - ethnicity (*p-value* < 0.0001)
 - gender (*p-value* = 0.0001)
- c statistic = 0.6

STUDY 2: MODEL 2

- of those who graduated, factors associated with graduating within 3 years include:
 - hours worked per week ($p\text{-value} < 0.0001$)
 - affiliated college upon entry to UCF ($p\text{-value} < 0.0001$)
 - ethnicity ($p\text{-value} = 0.031$)
 - financial need ($p\text{-value} = 0.042$)
- c statistic = 0.67

STUDY 2: KEY FINDINGS

- of the significant factors found to predict graduation, affiliated college upon entry had the largest estimated odds ratios(model 1)
 - students from the Hospitality Management, Health and Public Affairs and Nursing colleges are 3 times more likely to graduate than Engineering students
- of those who graduated, time to graduation was associated with the number of hours worked per week (model 2)
 - a student who reported working 11-20 hours per week was 2.5 times more likely to graduate while a student who reported working 21-39 hours was 1.5 times more likely to graduate within 3 years compared to a student who reported working 40 hours or more per week

IMPACT

- increase predictive power with models that contain IR and student survey variables
- implementing an ongoing business process to merge data sources:
 - positions us to proactively address strategic directions such as performance based funding metrics
 - helps us respond quickly for projects requiring higher order analysis

FUTURE WORK

- build predictive models using other survey sources
- model retention for specific student profiles of leavers
- drill into data
 - explore differential effects of predictive variables by specific student groups
 - dig into college variable – explore effect of performance in degree program gateway courses
- work with leadership to interpret findings and construct strategies to target specific student groups to increase success outcomes

CONTACT INFORMATION

Patrice Lancey

Assistant Vice President

patrice.lancey@ucf.edu

Uday Nair

Assistant Director

uday.nair@ucf.edu

Rachel Straney

Application Systems Analyst/Programmer

rstraney@ucf.edu

University of Central Florida

Office of Operational Excellence and Assessment Support

www.oeas.ucf.edu