Big Data: What’s the Big Deal?

Ellen D. Wagner Ph.D.
Chief Research and Strategy Officer
PAR Framework

@edwsonoma
edwsonoma@parframework.org
Common Definitions for Today

Data are bits of information, everywhere. They come in all kinds and shapes and sizes.

Analytics are methods and tools to parse streams of digital bits into meaningful patterns that can be explored to help stakeholders make more effective decisions.

Learning analytics are methods and tools to parse streams of digital bits into meaningful patterns that cognition, instruction and academic experience, including student success.

Data-readiness ranges from essential individual knowledge and skills to institutional capacity for creating a culture that values evidence-based decision-making.
Big Data Are Changing EVERYTHING
BIG DATA

VOLUME

VELOCITY

VARIETY

VERACITY
Google Trends

Topics
- big data

Interest over time

![Graph showing trends over time](image-url)
Google Trends

[Google Trends chart with topics such as big data, business intelligence, and learning analytics]

[Graph showing interest over time for selected topics]

[Logo and URL: parframework.org]
STAGGERING REVELATIONS ABOUT BIG DATA

Huge
2.8 ZBs of data have been created this year alone.

Baseline

http://bit.ly/1goTBmP
<table>
<thead>
<tr>
<th>Prefix</th>
<th>Abbreviation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigabyte</td>
<td>GiB</td>
<td>1,024 Megabytes</td>
</tr>
<tr>
<td>Terabyte</td>
<td>TiB</td>
<td>1,024 Gigabytes</td>
</tr>
<tr>
<td>Petabyte</td>
<td>PiB</td>
<td>1,024 Terabytes</td>
</tr>
<tr>
<td>Exabyte</td>
<td>EiB</td>
<td>1,024 Petabytes</td>
</tr>
<tr>
<td>Zettabyte</td>
<td>ZiB</td>
<td>1,024 Exabytes</td>
</tr>
<tr>
<td>Yottabyte</td>
<td>YiB</td>
<td>1,024 Zettabytes</td>
</tr>
</tbody>
</table>

1 ZB = 1,000,000,000,000,000,000,000,000,000 bytes
STAGGERING REVELATIONS ABOUT BIG DATA

Really Huge
The digital universe will reach 40 ZBs by 2020, which exceeds prior forecasts by 14%.

Baseline

http://bit.ly/1goTBmP
Really, Really Huge

40 ZBs is the equivalent of 1.7 MBs of new information created for every human on the planet for every second of the day.

http://bit.ly/1goTBmP
STAGGERING REVELATIONS ABOUT BIG DATA

Unexploited Opportunity

Only half of 1 percent (.5%) of the world’s data is analyzed, stressing the need for tech talent to extract value from this data.

http://bit.ly/1goTBmP
Where are we headed?

Business Models
Provide Guidance

Comprehensive Analytics
- KEY PERFORMANCE INDICATORS
- GEO-SEGMENTATION
- PERFORMANCE DASHBOARDS
- PATHING
- N-DIMENSIONAL ANALYSIS
- TRAFFIC PATTERNS
- EXCEL INTEGRATION

Online Business Optimization
- SEARCH ENGINE MARKETING
- CONTENT
- PROFILING AND BEHAVIORAL TARGETING
- CONVERSION FUNNEL
- A/B/C
- MULTI-CHANNEL ROI
- SCENARIO AND PROCESS
- CUSTOMER ACQUISITION

Courtesy Phil Ice, APUS
While “Big Data” raise expectations, student data drive big decisions in .edu
Costs and Completion Rates

Source: New York Times; NCES
Are You “Scorecard-Ready”?

http://collegecost.ed.gov/
Performance Based Funding and US Post-Secondary Institutions

Data Readiness in Higher Ed

Analytics have ramped up everyone’s expectations of personalization, accountability and transparency. Academic enterprises simply cannot live outside the institutional focus on tangible, measurable results driving IT, finance, recruitment and other mission critical concerns.
What do we want?
The RIGHT Answers!!

When do we want them? NOW!!
Use Case: The Predictive Analytics Reporting (PAR) Framework

• A national, non-profit, multi-institutional collaborative focused on institutional effectiveness and student success.
• A “massive data” analysis effort using using predictive analytics to identify drivers related to student risk
• PAR uses descriptive, inferential and predictive analyses to create benchmarks, institutional predictive models and to inventory, map and measure student success interventions that have direct positive impact on behaviors correlated with success.
PAR Framework per the Institute for Higher Education Policy

Predictive Analytics Reporting Framework
www.weti.wiche.edu/par

Launched in 2011, the Predictive Analytics Reporting (PAR) Framework is a non-profit multi-institutional data mining collaborative that brings together two year, four year, public, proprietary, traditional, and progressive institutions to voluntarily provide data and collaborate on identifying points of student loss and to find effective practices that improve student retention.

LEADERSHIP: Director: Beth Davis | Manager: Ellen Wagner

Number of Institutions by State

Participation

32 Institutions
14 States & Territories
1.8 million covered

*Includes institutions that are distance education only and may operate in more than one state

Draft, August 2014, used with permission, IHEP http://www.ihep.org/
PAR Framework

ACT

Predict

Retain

Reusable predictive models

Measurable intervention results

Common Definitions of terms

Common Definitions of interventions

Multi-Institutional collaboration

Scalable, cross-institutional improvements that support each individual student’s success
Structured, Readily Available Data

- Common data definitions = reusable predictive models and meaningful comparisons.
- Openly published via a cc license @ https://public.datacookbook.com/public/institutions/par
PAR Data Inputs

Student Demographics & Descriptive
- Gender
- Race
- Prior Credits
- Perm Res Zip Code
- HS Information
- Transfer GPA
- Student Type

Student Course Information
- Course Location
- Subject
- Course Number
- Section
- Start/End Dates
- Initial/Final Grade
- Delivery Mode
- Instructor Status
- Course Credit

Student Financial Information
- FAFSA on File – Date
- Pell Received/Awarded – Date

Student Academic Progress
- Current Major/CIP
- Earned Credential/CIP

Course Catalog
- Subject
- Course Number
- Subject Long
- Course Title
- Course Description
- Credit Range

Lookup Tables
- Credential Types Offered
- Course Enrollment Periods
- Student Types
- Instructor Status
- Delivery Modes
- Grade Codes
- Institution Characteristics

Possible Additional **
- Placement Tests
- NSC Information
- SES Information
- Satisfaction Surveys
- College Readiness Surveys
- Intervention Measures

** Future
## PAR Outputs

<table>
<thead>
<tr>
<th>Descriptive Benchmarks</th>
<th>Predictive Models</th>
<th>Intervention Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show how institutions compare to their peers in student outcomes, by scaling a <strong>multi-institutional database</strong> for benchmarking and research purposes.</td>
<td>Identify which students need assistance, by using <strong>in-depth, institutional specific predictive models</strong>. Models are unique to the needs and priorities of our member institutions based on their specific data.</td>
<td>Institutions address areas of weakness identified in benchmarks and models by scaling and leveraging a member, data and literature validated framework for examining interventions within and across institutions (<strong>SSMx</strong>)</td>
</tr>
</tbody>
</table>
5 Steps For Achieving (Learning) Analytics Success
START WITH AN EYE ON YOUR OUTCOMES.
BE CLEAR ABOUT WHAT YOU MEAN BY SUCCESS.
SHARED UNDERSTANDINGS ENABLE COMPARISONS, COLLABORATION.
FOCUS ON INSIGHTS, NOT JUST ON DATA.
SHARE YOUR WORK
For more information about PAR please visit our website:
http://parframework.org

Ellen Wagner:
Twitter http://twitter.com/edwsonoma
Google+ edwsonoma
On email ellen.wagner@parframework.org

THANK YOU FOR YOUR INTEREST