HOW I LEARNED TO STOP WORRYING AND TRUST LIVE DATA

Presented at the 50th Annual Meeting of the North East Association for Institutional Research, Baltimore, Maryland, November 6, 2023

Braden J. Hosch, PhD
Vice President for Educational and Institutional Effectiveness
Who are you going to be?
Dr. Strangelove

Presentation Outline

• Issues with live data
• Approaches to dial down the crazy and make it look more like magic
  • Organizational realignment
  • Data governance
  • Data management & handling

Dr. Strange
<table>
<thead>
<tr>
<th></th>
<th>25,865</th>
<th>1400</th>
<th>94</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2023 headcount enrollment</td>
<td>Fall data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median SAT (test optional)</td>
<td>Avg. high school GPA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>68% 32%</td>
<td>38% 30%</td>
<td>30% 21%</td>
</tr>
<tr>
<td>Undergrad Graduate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16,309</td>
<td>3,027</td>
<td>#58</td>
</tr>
<tr>
<td>Fall 2023 employees including hospital</td>
<td>Fall data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1957</td>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>USD annual budget 2023-24</td>
<td>Founded</td>
<td>Joined AAU</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How “live” is “live”?

- Instantaneous Quantum Computing
- Microseconds Transactional API
- Second to hours Periodic API/ETL
- Daily Traditional ETL
Why do we worry about live data?

IR Values and Data
- Accuracy & integrity
- External standards
- Replicability
- Meaning
- Trust

Issues with Live Data
- Transactional errors
- Internally defined
- Unrepeatable
- Uncertainty
- Unreliability
So what do institutional researchers do?

- **Take a “snapshot”**
  - Extract needed data at a specific point in time, curate it, and preserve it

- **Advantages**
  - Consistent, valid, reliable, auditable
  - Can always produce the same answer

- **Disadvantages**
  - Slow, out-of-date, limited access, labor-intensive
  - Backward-looking
Approaches to handle live data

- Organizational Realignment
- Data Governance
- Strategies for data management & handling
ORGANIZATIONAL REALIGNMENT

- Restructure
- Get rid of silos … by getting rid of silos
- Cultural challenges
Org Chart Detail

22.0 FTE direct
24.0 FTE w/ indirect

Not shown:
Dotted lines to Executive Director of CRAFT and Survey Research Analyst
## Institutional Research “vs” Business Intelligence
### Structural Differences

<table>
<thead>
<tr>
<th></th>
<th>Business Intelligence</th>
<th>Institutional Research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organizational Reporting</strong></td>
<td>Information Technology</td>
<td>Provost, Other VP area</td>
</tr>
<tr>
<td><strong>Data Realms</strong></td>
<td>Enterprise Systems</td>
<td>Data warehouse, curated data sets</td>
</tr>
<tr>
<td><strong>Constituencies</strong></td>
<td>Internal</td>
<td>Internal and external</td>
</tr>
<tr>
<td><strong>Data usage</strong></td>
<td>Operational decision support</td>
<td>Official reporting &amp; decision support</td>
</tr>
<tr>
<td><strong>Age of organization</strong></td>
<td>Newer</td>
<td>Well-established</td>
</tr>
</tbody>
</table>

# Institutional Research “vs” Business Intelligence

## Cultural Differences

<table>
<thead>
<tr>
<th></th>
<th>Business Intelligence</th>
<th>Institutional Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Educational Background</td>
<td>BA/BS, MS, IT related</td>
<td>PhD Common, Social Sci/Statistics</td>
</tr>
<tr>
<td>Career Background</td>
<td>Not academic, often not higher ed</td>
<td>Higher ed</td>
</tr>
<tr>
<td>Career Opportunities</td>
<td>Outside higher ed</td>
<td>Higher ed</td>
</tr>
<tr>
<td>Data tools</td>
<td>Provider perspective</td>
<td>User perspective</td>
</tr>
<tr>
<td>Data Manipulation</td>
<td>A crime</td>
<td>A necessity</td>
</tr>
<tr>
<td>Data Openness</td>
<td>Democratization of data</td>
<td>Manage carefully</td>
</tr>
<tr>
<td>Data Quality</td>
<td>Provenance fidelity to source</td>
<td>Conform to definition; fitness of use; consistency</td>
</tr>
<tr>
<td>Data understanding</td>
<td>Operational &amp; managerial context</td>
<td>Institutional &amp; strategic context</td>
</tr>
<tr>
<td>Attitude</td>
<td>Optimistic</td>
<td>Skeptical</td>
</tr>
</tbody>
</table>

DATA GOVERNANCE

• Structure
• Important features
Data Governance

Executive Sponsors

Data Governance Council

Data Stewards

Domain Steward
Domain Steward
Domain Steward
Domain Steward

Area Stewards

Data Governance Administration

Data Users

Source: Stony Brook Data Governance Council Structures
### Key features of data governance systems

<table>
<thead>
<tr>
<th>Documents</th>
<th>Groups</th>
<th>Individual roles</th>
</tr>
</thead>
</table>
| • Charter / framework  
  • Principles & values  
  • Purpose & scope  
  • Roles & responsibilities  
  • Written & published policies  
  • Data dictionaries  
  • Communication strategies | • Senior leadership/Executive sponsors  
  • Policy council  
  • Data steward council(s)  
  • Information security council/program  
  • Data Governance Office/Administrative Personnel | • Data stewards  
  • Data custodians/caretakers  
  • Data users |
DATA MANAGEMENT & STRATEGIES

• High frequency ETL
• Data reduction
• Manage/set alerts on values at error tolerance thresholds
• Load validation systems with alerts
Enrollment Comparatives

Frequency: daily

Initial problems: Transactional system dropped historical records and gave false impression of increase over historical performance.

(We thought we knew this data source, and we didn’t)

Solution: Daily thin snapshots. Took a year to execute.

Source: Stony Brook Analytics | Enrollment Comparatives Dashboard (by College/School)
Work order dashboards

Frequency:
Daily; weekly & monthly KPIs

Initial problems:
3rd party data source, import, Input accuracy

Solution:
Complex ETL, Employee training
Publicly available energy dashboards

Frequency: 15 min

Initial problems:
- 30 billion rows
- Whipsaws

Solution:
- Data reduction
- Limited metrics
- Tolerance ranges

Source: Stony Brook Energy Dashboards: https://energy.stonybrook.edu/dashboards/energy-map.html
Pulse survey dashboards

**Frequency:** Daily data feed, daily comments, weekly statistical update

**Initial problems:**
- Time period for statistical stability
- Coding comments

**Solution:**
- Detailed analysis over first six months
- Comment coding (pending-exploring AI)

---

Source: Stony Brook Analytics | Pulse Survey
Admissions Comparatives

Frequency: Daily

Initial problems:
Disagreement among users about refresh frequency, system of record

Solution:
ETL from source system instead of warehouse (remains unresolved)
What else do we need to make this work?

- Technology platform(s)
- Personnel
- Data source expertise
## Final Thoughts

<table>
<thead>
<tr>
<th>Live data is not a “set it and forget it” operation</th>
<th>Expect the unexpected</th>
<th>Understand your users</th>
<th>Using live data (or not) affects your brand</th>
</tr>
</thead>
</table>

- Live data is not a “set it and forget it” operation
- Expect the unexpected
- Understand your users
- Using live data (or not) affects your brand