

Beyond Data Governance to Data Strategy

Braden J. Hosch, Ph.D.

Asst. Vice President for Institutional Research, Planning & Effectiveness

Stony Brook University

AIR Annual Forum, Washington, DC, May 31

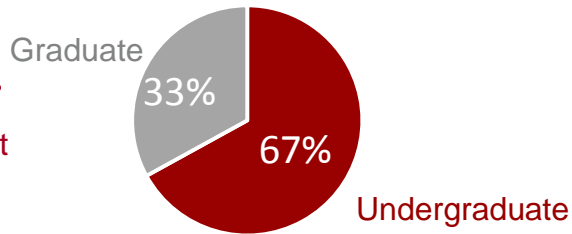
Overview

- Background
- What is a data strategy and why do we need one?
- Components of Stony Brook's data strategy
- Implementation (as process, not project!)

Institutional Profile

Students:

25,734
Fall headcount



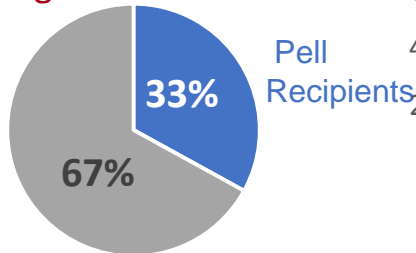
Institution:

Doctoral, Highest Research Activity
Public AAU
Founded 1957

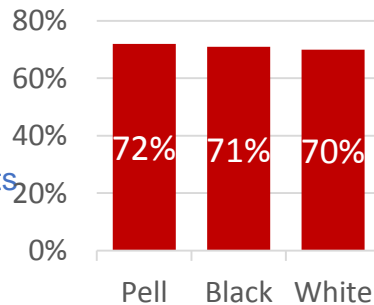
Undergraduate Profile

1254

avg. SAT

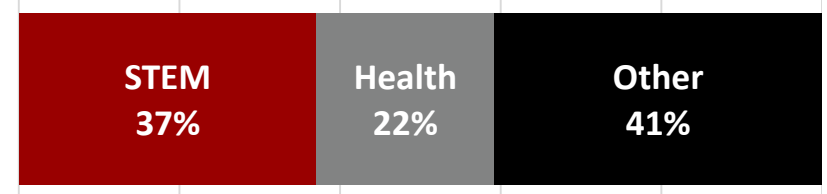


72% 6-yr grad rate



Program Profile

6,754 Completions 2015-16



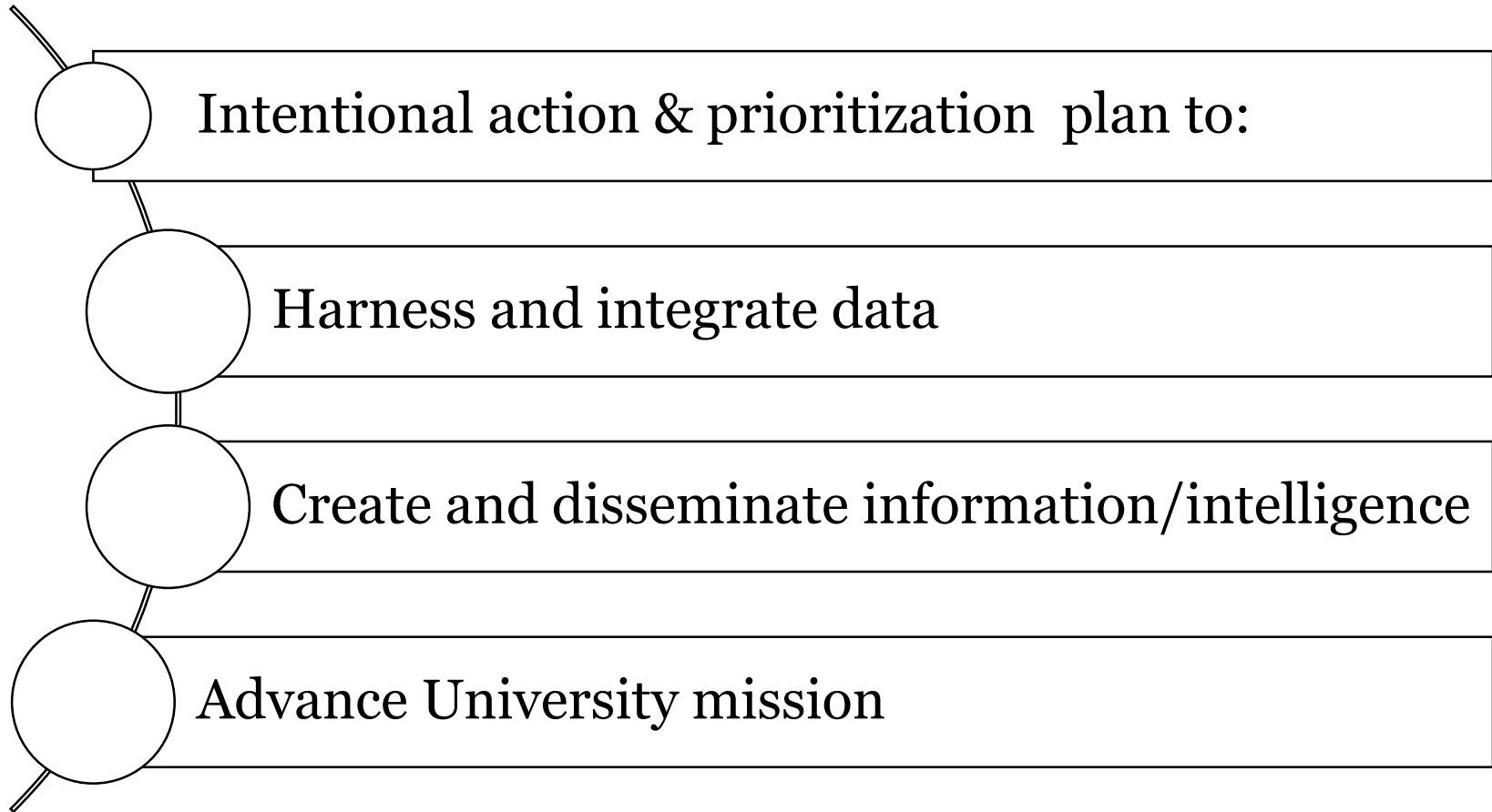
Employees:

14,732 including hospital
2,695 faculty (FT & PT)

Finance:

2.5 billion USD annual budget
230 million USD research exp.

What is a data strategy?



Why do we need a data strategy?

Support objectives to:

- Promote operational effectiveness, excellence & efficiency
- Retain and grow revenue
- Reduce risk
- Drive innovation

Proliferation of data assets

Increasing organizational size and complexity

Advances in analytical tools

Selected Stony Brook data assets



Assessment Data
Help Desk Tickets
Card Swipes
Surveys

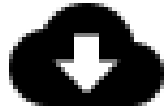


Stony Brook's mission

The university has a five-part [mission](#) to provide and carry-out:

- Highest quality comprehensive education
- Highest quality research and intellectual endeavors
- Leadership for economic growth, technology, and culture
- State-of-the-art innovative health care, with service to region and traditionally underserved populations
- Diversity and positioning Stony Brook in global community

Elements of Stony Brook's data strategy



Data acquisition



Data governance



Data quality



Data access



Data usage & literacy



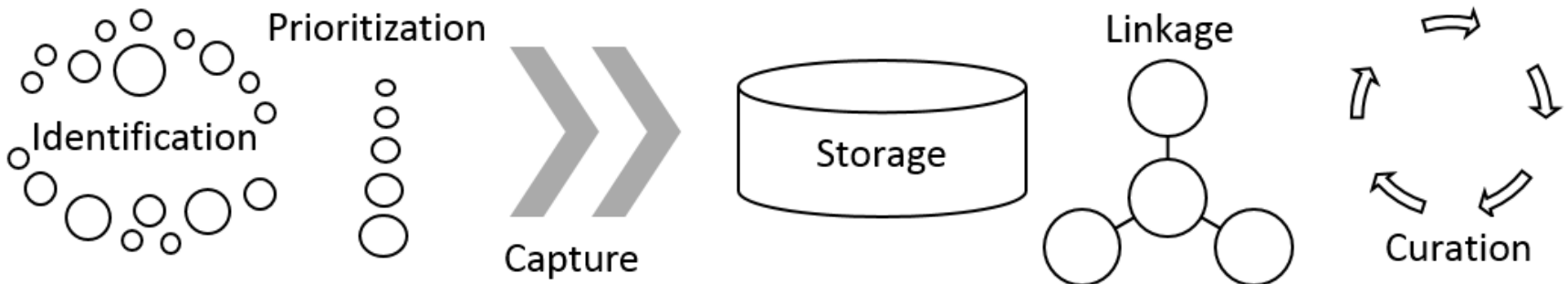
Data extraction & reporting



Data analytics

Data acquisition

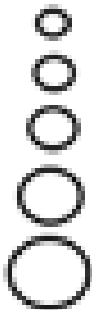
- Data acquisition involves identification, prioritization, capture, storage, linkage, and curation of data assets most valuable to the enterprise



Data acquisition Identification & prioritization



Prioritization

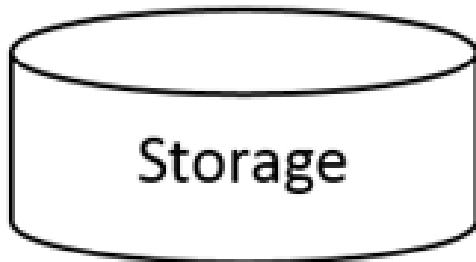


- Establish and maintain an inventory of data assets and assess acquisition maturity
- Establish a process to prioritize integration into data infrastructure

Data Acquisition Capture & storage



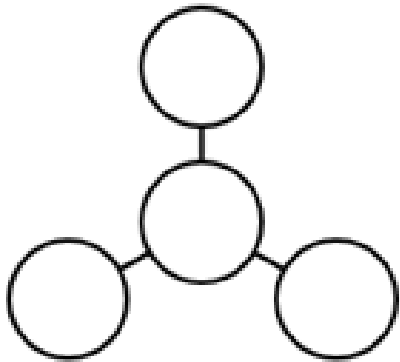
Capture



- For each data asset identify current and optimal capture procedures
- For each data asset identify current and optimal storage areas

Data Acquisition Linkage & curation

Linkage



Curation

- For each data asset identify current and optimal procedures to link to other data sources
- For each data asset identify how data will be updated and maintained to preserve value



Data governance

- Data governance formalizes behavior around how data are defined, produced, used, stored, and destroyed to enable and enhance organizational effectiveness.

PeopleSoft and the Data Warehouse are governed by the University Data Governance Council

Establish expectations for all other data assets to have formal data governance

Data governance Requirements

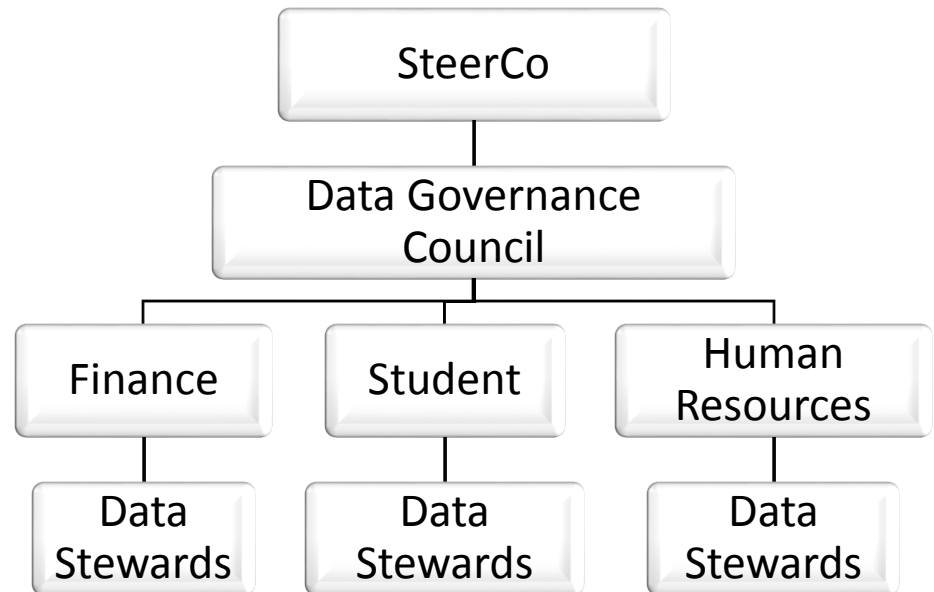
Designated decision-making body

Formal data dictionaries and descriptions of architecture

Individuals designated to provide stewardship

May opt to be governed through the Stony Brook Data Governance Council

Stony Brook Data Governance Framework*



*Applies to PeopleSoft and the Data Warehouse (as of 9/26/16)

Data Quality

- Data quality is the state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use.

The Data Governance Council is charged with improving data quality for PeopleSoft and the Data Warehouse. A roadmap to achieve this has been developed

For each data asset, develop and execute a plan to maintain and improve data quality; automate when justified by ROI.

Data access

- Data access ensures authorized individuals can obtain and use data when and where they are needed and protects privacy and sensitive information by preventing unauthorized use.



Accessibility



Authorization



Security



Data usage and literacy

- Data usage and literacy entail people regularly obtaining data; understanding them; and using them to improve operational effectiveness .

Establish for all data assets:

Usage metrics

Effectiveness metrics

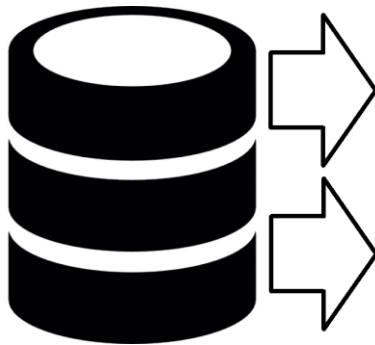
Training inventory

Data User Responsibilities

1. Recognize data complexities; understand data meanings and limitations
2. Cite sources; assume broad audiences
3. Respect privacy
4. Secure data and reports
5. Report data quality issues

Data extraction and reporting

- Data extraction and reporting represent the ways that data are queried and retrieved from storage and then delivered to users through regularized and ad hoc reporting to support day-to-day operations.



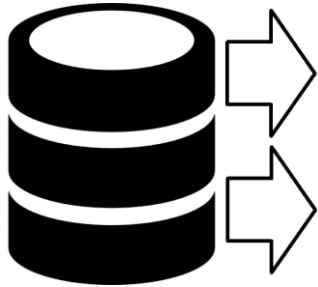
Extraction



Reporting



Data extraction and reporting



- Methods for querying and extracting data from storage should be identified, including user types associated with each extraction method

Extraction



Reporting

- Reports should be linked to operational objectives
- Report inventories should be maintained in an accessible area.
- Reports should be automated depending on ROI
- Reports should include effectiveness metrics



Data analytics

- Analytics deliver dynamic and visual analysis of data, internal & external benchmarking, exploratory and causal analysis, and predictive/forecasting capacity.

Requirements

Maturity in data acquisition, governance, quality, access, usage, & extraction

Tools capable of performing analyses and communicating effectively

Speed and ease of use

Data asset strategy document compiled for each data asset

Data Asset Strategy Doc
e.g. IPEDS

Description & use


 Data acquisition

Priority (High, medium, low)

	Current	Plan	Date
Capture			
Storage			
Linkage			
Curation			

 Data governance plan


 Data quality protocols

 Data access plan

Accessibility
Authorization
Security

 Data usage and literacy

 Data extraction/reporting

 Data analytics

Example – Description and Use, Priority

National Student Clearinghouse

Third-party repository of enrollments and some completions of post-secondary enrollments in participating higher education institutions in the United States. Data are used for monitoring subsequent enrollment of applicants, students leaving Stony Brook without degrees, and degree completers. Major types of data include:

- Institution of enrollment
- Institution characteristics
- Dates of enrollment
- Student level and enrollment intensity
- Award completion
- Field of award (later records)

Priority Level: Medium

Example – Data Acquisition

National Student Clearinghouse

	Current	Plan	Date
Capture	Special query required. Applicants uploaded and maintained by Enrollment Management; students transferring out and graduates queried and maintained by IRPE; Student ID reattached following query	Continue query protocol; standardize query dates	2017
Storage	Store on Enrollment Management file share, IRPE file share	Store in data warehouse SQL server	2017
Linkage	StudentID in Excel files – linkages are custom queries in IRPE	Housed in Data Warehouse with StudentID	2017
Curation	Applicant data maintained by Enrollment Management; other data cleaned through Access database to establish primary enrollment in fall term for grads and fall & spring terms for students transferring out	Write query into Data Warehouse; store analysis tables there	2017 and ongoing

Example – Data Governance, Quality

National Student Clearinghouse

Data governance plan

Once integrated into data warehouse, data would be under purview of Data Governance Council. Data Stewards: [Name removed], Enrollment Management (admitted undergraduates not enrolled), [Name removed], IRPE (all other data).

Data quality protocols

Raw data from the NSC are stored; annual updates are performed for students / graduates from 10 years prior through current year. Historical data are not overwritten even when updates appear in the NSC.

SQL in Access database uses a set of decision rules to identify a primary enrollment institution for fall and spring terms; IDs are unduplicated. Stony Brook enrollments are reattached from enrollment database regardless of NSC return file.

Example – Data Access Plan

National Student Clearinghouse

Accessibility – Student Tracker access limited to 2 enrollment management staff, and 3 IRPE staff. Return files accessible to two EM staff, and all IRPE staff. Future storage plan will extend access to 5 additional BI staff. Future access plan will extend analytics access to Academics Access Users and Executive Dashboard Users

Authorization – AVP of Enrollment management or AVP of IRPE required. Future authorization to access reports is TBD.

Security – Password authentication to NSC; password authentication to one of two university file shares. Note: **Institution name is a restricted data element** and may not be shared outside of Stony Brook. Future security will leverage Data Warehouse security protocols.

Example – Data Usage and Literacy

National Student Clearinghouse

Data Usage and Literacy

- **Dictionaries:** import NSC data dictionary for raw files, construct dictionary for cleaned files to explain eliminated records; integrate definitions with Tableau Server.
- **Documentation:** Construct 1-page of friendly documentation to accompany all reports discussing how data are collected, appropriate uses, and limitation.
- **Video:** This asset is not high priority for video training

Example – Extraction/Reporting National Student Clearinghouse

Extraction/Reporting

- Current reporting is ad hoc.
- Future reporting will provide regularized metrics for enrollment at other institutions and completion at other institutions for non-enrolled admitted undergraduates, transfer out undergraduates, completers at all levels.
- Unit record reports if developed would be delivered via SBU Reporting.

Example – Analytics

National Student Clearinghouse

Analytics

- No analytics are in place.
- Development of Tableau reports for Academics Users is primary deliverable.
- High level metrics may be developed for Executive Dashboard Users if requested.

Issues To Tackle

- Research data
- Prioritization process
- Resource allocation, with special attention to:
 - Storage
 - Security
 - User literacy

Wrap-Up and Discussion

- Creating and implementing a data strategy is a process, not an IT project → establish sustainable systems
- If it isn't written down and shared, then it's not a strategy, it's a secret
- This is just one approach. How are you considering a data strategy at your institution?