Fundamental Steps in Building an Effective Data Culture: Linking Planning, Ownership, Governance and Execution

Jason F. Simon, Ph.D.

Assistant Vice President – Data, Analytics, and Institutional Research

Introductions

▶ About Me

▶ Raise your hand if you:

- ▶ Work at a 2 year institution, 4 year institution, other entity or agency
- ▶ Data Analyst, Data Scientist, Assistant Director, Director, AVP, Other role
- ▶ Years in Higher Ed – 0-2, 3-6, 7-10, 10+

Agenda

▶ Why This Topic

▶ What is Data Culture?

▶ Culture-Centric Leadership: Your role in promoting a strong Data Culture

▶ Building Blocks and Best Practices (Executive Engagement to Data Governance)

▶ 10 Practical Things an IR Professional Can Do

▶ 7 Key Questions to Ponder Going Forward

▶ Q&A

Why This Topic?

Eduventures Releases 2017 Report on the New Higher Education Technology Landscape

January, 2018

Envisioning Pathways to 2030:

Megatrends shaping the future of global higher education and international student mobility

Is Your Institution Really Ready for Predictive Analytics?

By Darren Catalano   Jan 11, 2018

How Enrollment Challenges Can Spur Change

Amazon is quietly becoming its own university

Amy X. Wang   January 29, 2018

DIGITAL TRANSFORMATION
IN HIGHER EDUCATION

7 Ed Tech Trends to Watch in 2018

Competing in a world of sectors without borders

Your Students May Think Machines Teach Them More Than You Do

DIGITALIZATION
AND THE
AMERICAN
WORKFORCE

THE BIG DATA LANDSCAPE 2017

INFRASTRUCTURE

HADOOP ON-PREMISE
cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, bluedata, jethro

HADOOP IN THE CLOUD
amazon web services, Microsoft Azure, Google Cloud Platform, IBM InfoSphere BigInsights, TREASURE DATA, altiscale, Qubole, CenturyLink, CAZENA

STREAMING
amazon web services, databricks, confluent, striim, GridGain, METAMARKETS, DATATORRENT, data Artisans

NOSQL DATABASES
Google Cloud Platform, ORACLE, Microsoft Azure, MarkLogic, mongoDB, DATASTAX, HYPERTABLE, Couchbase, redislabs, influxdata

NEWSQL DATABASES
SAP, Clustrix, Pivotal, nuodb, Cockroach LABS, memsql, splice, VOLTDB, citusdata, TokuData, deepsql, paradigm4

GRAPH DBS
neo4j, Graph Cloud Platform, orient, kognitio, ArangoDB, Dremio

MPP DBS
TERADATA, VERTICA, NETEZZA, ORACLE, Pivotal, Infoworks

CLOUD EDW
amazon web services, Google Cloud Platform, Microsoft Azure, snowflake

DATA TRANSFORMATION
talend, pentaho, alteryx, TRIFACTA, tamr, Paxata, StreamSets

DATA INTEGRATION
informatica, MuleSoft, snapLogic, TEALIUM, Segment, enigma, podium data, alooma, xplenty, ZALONI, import.io, Stitch

DATA GOVERNANCE
informatica, APP DYNAMICS, actifio, WAVEFRONT, skyhigh, IBM, collibra, unravel, splunk, Alation

MGMT / MONITORING
amazon web services, New Relic, APP DYNAMICS, actifio, WAVEFRONT, BIAPps, unravel, splunk, Anodot, trocano, Numerify

STORAGE
amazon web services, Google Cloud Platform, Microsoft Azure, Cloudera, nimble storage, Qumulo, COHO, panasas

CLUSTER SERVICES
amazon web services, kubernetes, docker, MESOSPHERE, Core OS

APP DEV
rainforest

CROWDSOURCING
amazon mechanical turk, Upwork, WorkFusion, CrowdFlower

HARDWARE
Google TPU, ARM, Graphcore, MYTHIC, nvidia, Movidius, SCORTEX

ANALYTICS

DATA ANALYST PLATFORMS
Microsoft, pentaho, alteryx, Digital Reasoning, GUAVUS, AYASDI, ATTIVIO, Datameer, Quid, ClearStory, OrigamiLogic, interana, Bottlenose, ARIMO, ENDOR, MODE

DATA SCIENCE PLATFORMS
IBM, KNIME, dataiku, DOMINO, H2O, rapidminer, CONTINUUM ANALYTICS, Alpine, Angoss, ALGORITHMIA

BI PLATFORMS
Microsoft, DOMO, birst, GoodData

VISUALIZATION
Tableau, SAP, Google Cloud Platform, Qlik, looker, CELONIS, Periscope, ZoomData, CHARTIO, ZEPL, plotly

VERTICAL ANALYTICS
PREDIX, GE IoT, CAPE, UPTAKE, TACHYUS, Alluvium

STATISTICAL COMPUTING
SAS, SPSS, MATLAB

DATA SERVICES
Palantir, OPERA, DATA SCIENCE, kaggle, EXL, DataKind, FF

MACHINE LEARNING
H2O.ai, DataRobot, SKYTREE, bonsai, dataRPM, deepsense, Domino, CURIOUS AI

HORIZONTAL AI
IBM Watson, Cortana, Face++, sentient, Voyager, skymind, darifai, Affectiva, euroconom, nara, OSARO, BLUE VISION

SPEECH & NLP
Google Cloud Platform, twilio, amazon, Narrative Science, semantic machines, NUANCE, x.ai, find, ARRIA, TalkIQ, IDIBON, snips, Gridspace, Vicarious

SEARCH
elastic, ORACLE, ThoughtSpot, Algolia, Lucidworks, swiftype, MAANA, alphasense, SINEQUA

LOG ANALYTICS
splunk, sumologic, loggly, kibana, logz.io

SOCIAL ANALYTICS
Hootsuite, NETBASE, DATASIFT, tracx, synthesio, onyx, bitly, predata

WEB / MOBILE / COMMERCE ANALYTICS
Google Analytics, AMPLITUDE, mixpanel, sumall, Airtable, retention, SIGOPT, granify, custora

APPLICATIONS – ENTERPRISE

SALES
CHORUS, INSIDESALES.COM, conversica, clari, AVISO, TACT, fusemachines, TROOPS, Datahug, ZENGAGIO

MARKETING - B2B
RADIUS, App Annie, EVERSTRING, Lattice, infer, MINTIGO, 6sense, tubular, Radius, Datahug

MARKETING - B2C
Zeta, bloomreach, blue yonder, persado, ACTIONIQ, kahuna, BLUECORE, SAILTHRU, QUANTIFIND, mparticle, Amplero, msg.ai

CUSTOMER SERVICE
MEDALLIA, zendesk, Gainsight, CLARABRIDGE, CLICKFOX, NG DATA, DigitalGenius, AUTOMAT, frame.ai

HUMAN CAPITAL
HireVue, entelo, hiQ, textio, Wade&Wendy, Cushieo, Stella, pymetrics

LEGAL
RAVEL, Seal, GIGSTER, Brevia, ROSS

FINANCE
anaplan, Zuora, tidemark

ENTERPRISE PRODUCTIVITY
slack, X.ai, ORACLE

BACK OFFICE AUTOMATION
Captricity, Cora, talla, butter.ai, KASISTO

SECURITY
TANIUM, CYLANCE, StackPath, illumio, CODE42, DARKTRACE, ThreatMetrix, DataGravity, VECTRA, CyberCloud, Guardian Analytics, ANOMALI, sift science, CONFYD, cloudera, SecurityScorecard, Elastica, Recorded Future, feedzai

APPLICATIONS – INDUSTRY

ADVERTISING
AppNexus, MediaMath, criteo, xAd, Integral, MOAT, theTradeDesk, OpenX, dstillery, drawbridge, TAPAD, DataXu, gumgum, Quantcast, Rocketfuel

EDUCATION
Clever, declara, kidaptive, Panorama, knewton, gradescope

GOVERNMENT
Socrata, OPENGOV, OpenDataSoft

FINANCE - LENDING
OnDeck, Affirm, Kreditech, AVANT, INSIKT, MoneyLion, enigma, TrueAccord, blooom

FINANCE - INVESTING
Dataminr, KENSHO, Quantopian, NUMERAI, cignifi, aire, clarity money, Algoriz

REAL ESTATE
Opendoor, VTS, Quantopian, CREXi, COMPSTAK

INSURANCE
metromile, Lemonade, CYENCE, Tractable

HEALTHCARE
FLATIRON, Health IQ, Gingerio, Glow, COTA, AiCure, imagia

LIFE SCIENCES
KYRUUS, color, zymergen, BenevolentAI, zebra, Atomwise

TRANSPORTATION
UBER, ZOOX, TESLA, nauto, drive.ai, nexar, comma.ai, Civil Maps

AGRICULTURE
TRACE, farmersEdge, CLEARPATH, FarmLogs, BLUE RIVER, mavrx, Descartes Labs

COMMERCE
Instacart, RetailNext

OTHER
stem, HowGood, celect

CROSS-INFRASTRUCTURE/ANALYTICS
amazon web services, Google Cloud Platform, Microsoft, IBM, SAP, Hewlett Packard Enterprise, SAS, 1010data, vmware, TIBCO, TERADATA, ORACLE, NetApp

OPEN SOURCE

FRAMEWORK
Hadoop, Flink, Spark, MESOS, YARN, TEZ, CDAP

QUERY / DATA FLOW
Spark SQL, presto, cassandra, SLAMDATA, Google Cloud Dataflow

DATA ACCESS
nifi, mongoDB, cassandra, RSQDB, CouchDB, riak, HBASE, Spanner, ACCUMULO

COORDINATION
talend, Apache Zookeeper, Apache Ambari

STREAMING
Spark, APEX, Flink, kafka, druid, STORM

STAT TOOLS
R, ScalaLab, SciPy

AI / MACHINE LEARNING / DEEP LEARNING
TensorFlow, Caffe, CNTK, DMTK, Keras, VELES, WEKA, neon, FeatureFu, DIMSUM, DSSTNE, mllib, Aerosolve

SEARCH
elasticsearch, Solr, Lucene

LOG ANALYSIS
elasticsearch, kibana, logstash

VISUALIZATION
BEAKER, Rodeo

COLLABORATION
Jupyter, ANACONDA

SECURITY
Apache Ranger, KNOX, Sentry

DATA SOURCES & APIS

HEALTH
JAWBONE, VALIDIC, practicefusion, fitbit, GARMIN, Human API, kinsa

IOT
GE Digital, UPTAKE, ThingWorx, helium, samsara

FINANCIAL & ECONOMIC DATA
Bloomberg, THOMSON REUTERS, DOW JONES, S&P CAPITAL IQ, CB Insights, xignite, quandl, YODLEE, PREMISE, estimize, SECOND MEASURE, Eagle Alpha, StockTwits, PLAID

AIR / SPACE / SEA
PLANET, Airware, spire, AEROBOTICS, SPACEKNOW, Drone Deploy

PEOPLE / ENTITIES
acxiom, Experian, EPSILON, InsideView, Crimson Hexagon, SAFEGRAPH, BASIS

LOCATION INTELLIGENCE
foursquare, Sense, PlaceIQ, esri, factual, CARTO, Mapillary, panjiva

OTHER
qualtrics, DATA.GOV, dataworld, STREETLINE

DATA RESOURCES

INCUBATORS & SCHOOLS
PLURALSIGHT, galvanize, DataCamp, DataCite, INSIGHT

RESEARCH
facebook research, OpenAI, MIRI, DataKind, DRIVENDATA, Ai2 ALLEN INSTITUTE FOR ARTIFICIAL INTELLIGENCE

Our Data Landscape is Changing

▶ The velocity, volume, and speed of data is crushing

▶ Most institutions have not yet fully realized the analytic potential of a robust data landscape (Bichel, 2012).

▶ Said differently Reinetz (2015) stated that "higher education is data rich but information poor (p.4)."

▶ Future growth in data competency is predicated on leveraging institutional data differently

▶ Institutions and IR need to approach their data in new ways

▶ Reporting of official, often static, information as the norm is no longer good enough for modern higher education institutions

▶ True value from a data landscape is when the institution can leverage existing data to answer problems focused on the future NOT the past

How Might This Change Impact IR?

- We are being pushed to improve our outcomes

- The competitive marketplace is evolving and applying pressure...everywhere from Academic Affairs to Student Life

- Campuses may be looking for quick fixes for data challenges

- Focusing on tools and technology alone is not enough

- Institutions of higher education are made up of faculty, staff, students, and alumni.

- These groups all contribute to an organization's data culture and influence prioritization activities.

So What is Culture?

▶ The customary beliefs, social forms, and material traits of a social group

▶ The characteristic features of everyday existence (such as diversions or a way of life) shared by people in a place or time

▶ The set of shared attitudes, values, goals, and practices that characterizes an institution or organization

▶ The integrated pattern of human knowledge, belief, and behavior that depends upon the capacity for learning and transmitting knowledge to succeeding generations

Merriam-Webster Dictionary

How would you categorize our IR Data Culture?

- ▶ Customary beliefs - ?

- ▶ Social forms - ?

- ▶ Material traits - ?

- ▶ Features of everyday existence - ?

- ▶ Attitudes - ?

- ▶ Values - ?

- ▶ Goals - ?

- ▶ Practices - ?

- ▶ Capacity for learning - ?

- ▶ Transmitting Knowledge- ?

Adapted from Merriam-Webster Dictionary

How Others MIGHT See IR Data Culture

► Customary beliefs – Official Data, Rigid Data Definitions, and Autopsy Focused

► Social forms – Nice conversations with Analysts/Directors, committee service, etc.

► Material traits – Factbooks, Output, Reports, Numbers People

► Features of everyday existence – Ad Hocs, Paper Reports, National Surveys, Rankings

► Attitudes – Good at what they do, seem stressed, busy, responsive, slow to respond

► Values – Accuracy, Integrity of Information, Perfectionists, Rather get it right then get it done quick

► Goals – Be responsive to my needs, Be able to provide me answers

► Practices – coding, manipulating, tabulating, categorizing, cleaning, etc.

► Capacity for learning - eager but not much time to do it beneath all the ad-hocs

► Transmitting knowledge – would rather pass on data to get to the next ad-hoc then explain it to me in a way that might help inform my next steps (or the opposite)

Adapted from Merriam-Webster Dictionary

If Your IR Shop Wants to Evolve What Might Get in Your Way?

- Data initiatives and changing culture can get sidetracked by:
  - Prioritization disagreements
  - Data ownership conflicts
  - Turf wars
  - Poor resourcing
  - Lack of executive support
  - Confusion over data responsibility
  - A lack of formalized roles and responsibilities around data governance and management
  - Resistance to change out of fear
  - Other Thoughts?

The Case for a Culture-Centric IR Leader

- Confronting these issues requires:
  - Planning, effort, and a conscious decision to for the institution to reflect on current behaviors and norms.
  - Paying attention to the people, processes, programs, and spoken and unspoken rules around a given data project in a given environment (department, division, campus, university system, etc.).

- Culture-centric data leaders who:
  - Recognize the human element of their work
  - Approach data prioritization with a different set of assumptions and practices
  - Value the role of people, processes, and structures.

So How Do We Go About Doing This?

Start at the Top

- Identify and Engage an Executive Champion
  - The time for enthusiastic support from senior leadership within colleges and universities is more important than ever.
  - Organizations that develop plans to solicit early and continued support for data projects with champions reap the benefit of the investment of time in this effort.
  - Succeeding in the data marketplace today requires executive champions with specific goals in mind for their institutions.
  - This is important especially if your culture is in opposition to these expectations.
  - Ultimately, an executive champion is very helpful in influencing culture change on items that have historically been fuzzy or unclear to the wider institutional community.
  - **This may require IR leadership to seek out access and time – you may need to build consensus upward through your chain of command.**
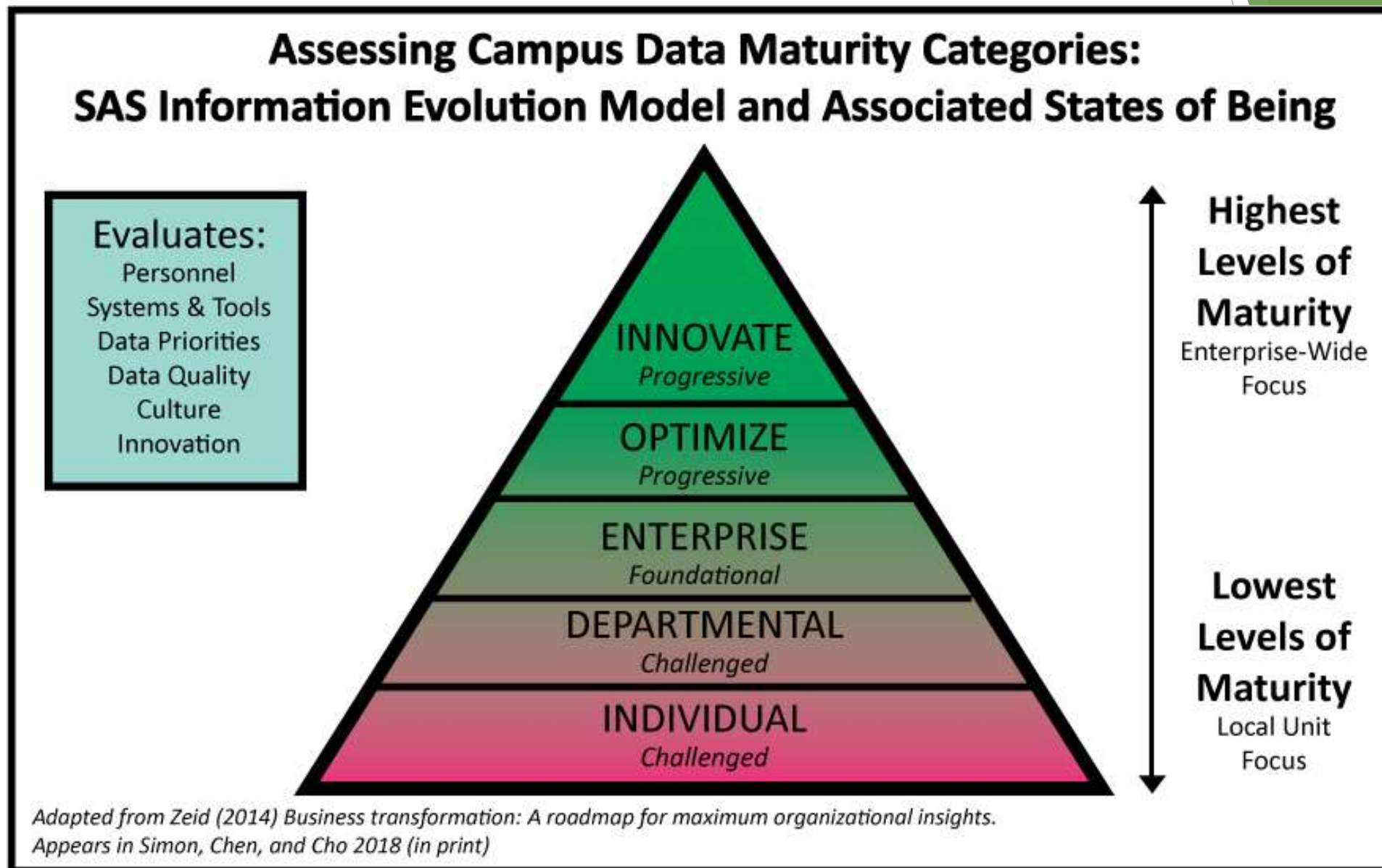
Conduct a Data Maturity and Analytics Climate Audit

- Prioritize time to assess the larger institutional data culture is a key step

- Achieving this goal requires substantive conversations about the maturity of an institution's campus culture when it comes to data and increasingly the leveraging of data for analytic purposes.

- Data maturity is the capacity for an organization or campus to achieve maximum benefit from data related assets. These assets will either hamper or accelerate outcomes from prioritization activities.

- Zeid (2014) asserts that these assets include people, processes, technical infrastructure, and culture. In a mature organization these assets align to achieve maximum analytic efficiency, discover innovative solutions to common organizational challenges, and deliver better solutions and services to constituents.

Data Maturity : Where is your campus?

Assessing Campus Data Maturity Categories:
SAS Information Evolution Model and Associated States of Being

Evaluates:
Personnel
Systems & Tools
Data Priorities
Data Quality
Culture
Innovation

INNOVATE
Progressive

OPTIMIZE
Progressive

ENTERPRISE
Foundational

DEPARTMENTAL
Challenged

INDIVIDUAL
Challenged

Highest Levels of Maturity
Enterprise-Wide Focus

Lowest Levels of Maturity
Local Unit Focus

Adapted from Zeid (2014) Business transformation: A roadmap for maximum organizational insights.
Appears in Simon, Chen, and Cho 2018 (in print)

Data Maturity : Where is your campus?

- ▶ Take 5 minutes, and turn to your neighbor.

- ▶ Share where you think your campus is on the maturity index?

- ▶ Be prepared to share back with the larger group.

Other Tools for Assessing Data Maturity

▶ Educause has surveys you can download and complete for free

▶ Your campus may have participated in the Faculty ECAR study on technology

▶ Your IT department may have future-focused strategic planning documents (investment roadmaps)

▶ Your fellow colleagues from TAIR may be able to describe their campus tools and maturity – plan a visit to more robust campuses

▶ You could ask about data governance practices (more on that later)

Stakeholder Identification is Key

▶ Important consideration when moving from overall institutional data prioritization to individual subject matter areas.

▶ Seek out those responsible or accountable for data and colleagues who are often consulted or informed about data developments. This process is formally known as a RACI exercise.

▶ Leverage existing source documents (IT charters, taskforce findings, etc.) on campus to identify subject matter experts (SMEs) or technical SMEs.

▶ Recognize that the college or university community will not react homogenously to new technology.

▶ Ensuring that these voices are present in planning processes is an important step in ensuring diverse stakeholder opinions are heard.
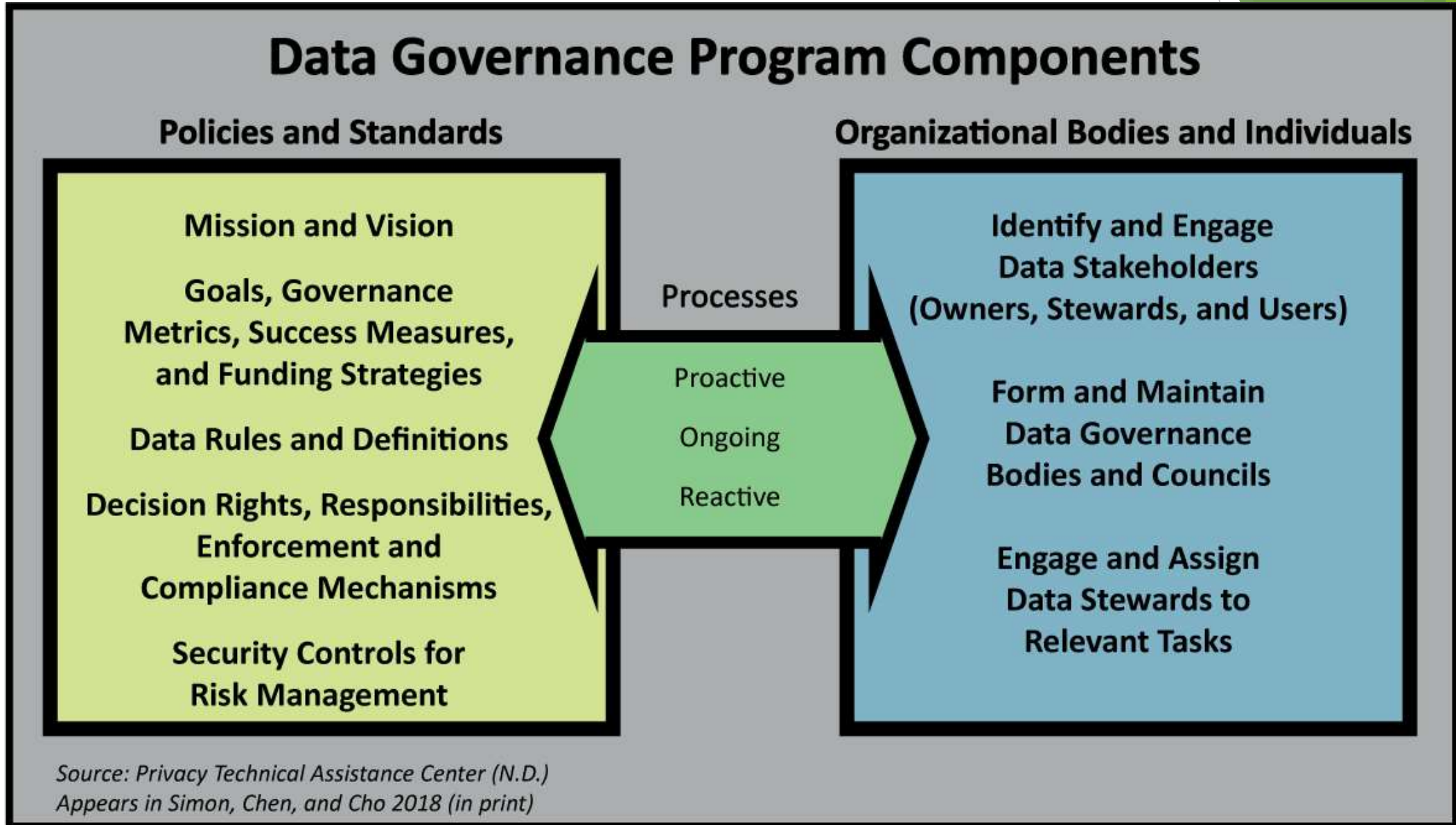
Effective Data Governance (DG) Impacts Data Culture - The BIG Challenge

▶ The Privacy Technical Assistance Center defines data governance "as an organizational approach to data and information management that is formalized as a set of policies and procedures that encompass the full life cycle of data, from acquisition to use to disposal" (pp.1-2).

▶ IR must recognize that there is no longer a simple way to control data or to ensure a single version of the truth exists on a campus where numerous data stakeholders may utilize a wide range of datasets and tools (Swing and Ross, 2016).

▶ Investing time in learning more about data governance is not only important but increasingly strategically necessary.

▶ DG Forces institutions to confront unspoken norms, tackle ownership issues, and define the operating principles for how various data users throughout campus will need to behave with one another.

Data Governance Explained in One Slide

Data Governance Program Components

Policies and Standards

Organizational Bodies and Individuals

Mission and Vision

Goals, Governance Metrics, Success Measures, and Funding Strategies

Data Rules and Definitions

Decision Rights, Responsibilities, Enforcement and Compliance Mechanisms

Security Controls for Risk Management

Processes

Proactive

Ongoing

Reactive

Identify and Engage Data Stakeholders (Owners, Stewards, and Users)

Form and Maintain Data Governance Bodies and Councils

Engage and Assign Data Stewards to Relevant Tasks

Source: Privacy Technical Assistance Center (N.D.)
Appears in Simon, Chen, and Cho 2018 (in print)

10 Practical Steps an IR Leader Can Take (1)

1. Read as much as you can about differences in organizational culture between the various divisions of a higher education institution. Recognize that each division will have its own set of expectations, requirements, and needs from data and data tools.

2. Investigate if a data maturity audit has occurred or if you need to consider starting a process.

3. Review old IT project charters and whitepapers to identify possible stakeholders, data pitfalls, and prioritization challenges from the past.

4. Start with a lunch. Gather like-minded data colleagues from around campus to begin conversations around the ideal state of data on your campus. Develop some next steps to expand your circle of influencers.

5. Review executive sponsor concerns raised in press releases, internal communications, or formal requests Institutional Research, Business Intelligence Unit, or Information Technology to understand opportunities for engagement.

10 Practical Steps an IR Leader Can Take (2)

6. Examine peer campuses – where are they in their data governance efforts? Consider site visits to learn more and see different structures in action.

7. Take a course in story-telling. Connect the seemingly disparate roles of data leader with story teller to advance your organization through data prioritization activities.

8. Conduct a review of data policies and procedures. Identify gaps and develop plans to partner with relevant campus units to address.

9. Consider stakeholder focus groups, surveys, or other feedback gathering opportunities to build your understanding of the campus data culture.

10. Hold a data summit. Provide the structure and the agenda but then listen...carefully.

Reflection Questions: The Take-A-Ways

- How would campus stakeholders describe the campus culture related to data?

- Where does your campus fit in terms of data maturity and practice?

- What strategies will you put in place to ensure that key constituents and stakeholders are effectively engaged in ways that are consistent with your campus culture?

- What are some strategies you would utilize to engage an executive sponsor? How have previous projects engaged these individuals? What mechanisms would you put in place to encourage and foster his/her support through this process?

- What data systems exist on your campus and where would individuals be categorized on a RACI matrix for each system?

- How might you leverage data governance practices to improve the data prioritization and data quality of your campus?

Discussion?

CONTACT INFORMATION:
Jason can be reached at Jason.simon@unt.edu or
940.565.2085

References:

- Inman, M. (2018). RACI Explained by Star Wars. Retrieved from: https://www.linkedin.com/pulse/raci-responsibility-model-explained-star-wars-matthew-inman/ on February 4, 2018.

- Privacy Technical Assistance Center (n.d.). *Data governance and stewardship.* Retrieved from: https://nces.ed.gov/programs/ptac/pdf/issue-brief- data-governance- and-stewardship.pdf

- Simon, J.F., Chen, P. and Cho, A. (In Print). Fundamental Steps in Building an Effective Data Culture: Linking Planning, Ownership, Governance and Execution. Routledge. New York, NY.

- Swing, R. L. and Ross, L. E. (2016). A new vision for institutional research. *Change: The Magazine of Higher Learning, 48*(2), 6-13.

- Zeid, A. (2014). *Business transformation: A roadmap for maximizing organizational insights*. Hoboken, NJ: John Wiley and Sons.