STAFFING FOR DATA SCIENCE IN FEDERAL AGENCIES

DR. MICHAEL VALIVULLAH
CTO, NASS-USDA

FCSM CONFERENCE - DECEMBER 6, 2016
DATA SCIENCE - MAJOR DOMAINS

- Business Expertise
- Analytical Expertise
- Data & Systems Expertise

Ref: Booz Allen Hamilton, NIST and McKinsey Global Research
DATA SCIENTIST - DEFINITION

‘A data scientist is a practitioner who has sufficient knowledge in the overlapping regimes of business needs, domain knowledge, analytical skills, and software and system engineering to manage the end-to-end data processes through each stage in the data lifecycle.’ - NIST
DATA SCIENCE – A TEAM SPORT

As there is a huge skills deficit of big data analytics professionals with:

Practical, Real-world Industry, Business and Technology Experience

Often experts from individual fields above with overlapping familiarity and interest in other fields are tapped to join a data science team.

This usually results in a multidimensional view for: describing, understanding, analyzing and solving data problems.
DATA SCIENCE TEAM ~ BACKGROUNDS

Who is on a data science team?

Advanced Masters/PhD degrees in:

DATA SCIENCE – POSITIONS & ROLES

- Data Officers
- Data Scientists
- Data Miners
- Data Modelers
- Data Visualization Designers
- Data Lake Modelers, Hadoop Engineers
- Data Analysts, ETL Developers, Cloud Experts
- Predictive Analysts
- Quantitative Analysts
- Big Data Architects & Data Engineers
DATA SCIENCE - TOOLS

Some Databases used:
• Oracle / Exadata / SQL Server / Teradata
• Cassandra / Hadoop / MapReduce / HBase
• Aster / Greenplum / Netezza / InMemory / Vertica

Some Languages used:
• C / C++ / CSS / HTML5 / Java / JavaScript / Perl / Python / Scala
• Hive / Pig / Lucene / Mahout / Solr / NoSQL / SQL

Some Statistics & Forecasting tools used:
• Angoss / MATLAB / R / SAS / SPSS / Watson
• ARCH / GARCH / SVAR / VAR / VEC / GAUSS
Some Data Visualization tools:
• QlikView / Spotfire / Tableau / yWorks

Some BI & Reporting tools:
• BusinessObjects / Hyperion / Cognos / MicroStrategy / OBIEE
BUILDING - ANALYTICS DRIVEN CULTURE

How to build an Analytics Driven Culture?

1. **Visualize Outcomes** - Identify business problems & desired outcomes
2. **Include Experts** - In industry/domain, people familiar with data, analytics, operations & computation
3. **Generate Insights** - Develop, test and evaluate hypotheses and hunches
4. **Realize Value** - In operations, organization and culture
1. Be Inclusive – Data Science is more than Data Scientists
2. Provide a way into the club – People learn better by doing
3. Flatten Hierarchy – Ivory Tower is the enemy of collaboration
4. Democratize Analytics – Everyone needs to be conversant
5. Rethink tolerance for Failure – It is the price of Discovery
6. Measure and reward - The ‘means’ as well as the ‘end’
CURRENT ORGANIZATIONAL STATUS

1) Willing Organization:
Believes in the power of analytics but do not know how to begin

Status: Beginner

How to move forward?

• Begin with a vision
• Explore available data to help inform the organization’s vision for analytics
• Chart a step-by-step path to achieve that vision
CURRENT ORGANIZATIONAL STATUS

2) Hesitant Adopter:

Analytics team is motivated to invest more in analytics but lacks organizational support

Status: Beginner/Middle of Journey

How to move forward?

• Build a Prototype to dispel doubt
• Use prototypes to prove value of analytics
• Think big but start small
• Work with end users and help overcome doubt
CURRENT ORGANIZATIONAL STATUS

3) Data Distressed:

Wants to move towards sophisticated analytics but unable to get underlying data in order (data-disorder/disaster)

Status: Beginner/Middle of Journey

How to move forward?

• Bring stakeholders together
• Build bridges between IT and analytics teams
• So that data can be efficiently used by both
CURRENT ORGANIZATIONAL STATUS

4) Scaling-Up:

Analytics added value in some business areas but not all
Leaders are not sure how to expand and grow capabilities

Status: Middle/Far along the Journey

How to move forward?

• Set up a dedicated team (Eg. Center of Excellence) to foster collaboration, share best practices, lessons learned, standards and governance
CURRENT ORGANIZATIONAL STATUS

5) Early Adopter (almost there but):
Analytics generate insights, but not everyone has adopted analytics to drive decisions. They do not use analytics to answer new questions

Status: Far along the Journey

How to move forward?

• Initiate Culture Change
• Make analytic models easier to use and transparent
• Help end-users understand the models and outputs to drive full adoption
1. **Centralized Model:**

Business Units bring problems to a centralized Data Science Team overseen by Chief Data Scientist.

2. **Diffused Model:**

Data Science teams fully embedded in business units and report directly to individual business unit leaders.

3. **Deployed Model:** (Reverse of centralized model)

Data science teams are fully allocated to business units but overseen by Chief Data Scientist.
HOW TO BUILD A PROTOTYPE

1. Choose a knotty problem to solve
   Caution: Do not pick a real easy one Or an impossible one

2. Shoot for high return or high impact output
   Demonstrate business value.
   Analytics outputs should be actionable.
   People want to see results before buy-in.
3. Pick an iterative solution. Quick success is important. Solution need not be perfect but needs to be ‘real’ to gain buy-in and it needs to be actionable.

Caution: Do not pick a problem requiring extensive data collection and complex analytics that is hard to explain.

4. Try for outputs beneficial to all/several business units.

Positive results for only a small portion of stakeholders will limit organizational buy-in or adoption will only be in pockets.
CHIEF DATA SCIENTIST/OFFICER

CDO/CDS is a Data Evangelist (60%) and an Enforcer (40%)

1. Builds a Data driven and Analytics Culture (Visionary and Tactical)

2. Effective collaborator amongst all layers of the organization with exceptional presentation and communication skills; understands the organizational culture and office politics

3. Conversant in business use cases, data usage patterns, data sources across the organization, external touch points, business processes, best practices, operations,

4. Good knowledge of data modeling, data standards, systems engineering, analytical tools, data warehousing, computer languages, development tools, Agile Devt., DevOps., Cloud, Configuration Management, etc.

5. Experienced in policy devt., data governance, compliance, risk management (GCR); confidentiality, integrity and availability (CIA).

DATA SCIENCE – TALENT MANAGEMENT MODEL

1. Job Analysis: Activities and Tasks to be performed in a given job, role, position

2. Competency Framework: KSAs for successful job performance

3. Position Description: Define and differentiate primary job responsibilities

4. Hiring Criteria: Candidate ability, tech skills, personality traits for success

5. Workforce Modeling: Type and Amt of work with respect to available talent supply and demand
DATA SCIENCE – TALENT MGMT. MODEL - CONTD.

6. Workforce Design: Composition of workforce, how work should be distributed across existing capabilities

7. Career Path: Trajectory and Transition Points for personnel

8. Training and Course Devt.: Build new capability among junior talent and strengthen and expanding existing skillset among senior talent

9. Retention Strategy: Reward structure and work environment that is nurturing, inclusive and respectful

10. Organization and Team Building: Further develop leadership skills and team dynamics
SUCCESSFUL TRAITS

• Commonsense
• Team Spirit
• Curiosity/Creativity
• Perseverance/Resilience
• Visualization/Good Storytelling Ability
THE END

Thank You!

Michael.Valivullah@NASS.USDA.gov