Draining the Data Swamp

Doug Hague
dhague@outlook.com

Data, data everywhere without an insight to see*

*With apologies to Samuel Taylor Coleridge
Move From a Legacy Morass to Analytical Competitor

1. Speak the language of business

2. Start with what you have

3. Fund your growth through business growth
Speak the Language of Business

• Ensure *understanding* or be *ignored*
  • The need for *storytelling*
  • Use of anecdote
    • BUT statistically valid anecdotes not outliers which is what leaders will usually bring up

• *Trust* is critical
  • Analytics will be challenged by business leaders
  • Know the current business lore
    • Confirm or challenge it with analytics results
  • *Arrogance* is punished with apathy or worse

• Utilize analytically-minded *talent in the business*
  • Decision makers have people they already trust: Make them champions
Start with What You Have

• Data
  • Plenty of data exists in your corporation
  • Spreadsheets, Finance, reporting, operational systems
  • Get it in your hands and put it on a single disk
  • Ability to blend will be a gap

• Talent
  • There is talent in your corporation
  • Decision makers today demand data
    • A few know where it is and can put their hands on it: FIND THEM!
  • Storytelling is likely a big gap

• Systems
  • Operational systems likely collect data electronically
  • There may have been several data warehousing efforts-there may be a gem!
  • Don’t forget the spreadsheets-this may be some of your most important data
Do Everything From The Start

• If you wait, you lose
• Build gradually, but have them all
• Don’t forget the base needs

*Adapted from Health Analytics by Jason Burke
Data Does **NOT** Need to be Perfect

- Data quality is a **journey**
  - Set up the structure to measure and then continuously improve
  - Set up a data governance policy and program
  - Be sure to know where your data is going
- Data Scientists will always want more data
  - Budget data
  - Sell modeling
- Insights can be gained from **imperfect data**
  - Data has never been perfect, yet analytics teams have succeeded!
    - ALWAYS know the limitations and be sure to communicate them
Reporting-The Necessary Evil of Analytics

• Reporting is a **requirement** of all entities
  • It doesn’t have to be done in Analytics
  • If not done well, reporting will become Analytics responsibility and use resources

• Some of the **best ideas** and questions will come because of report requests
  • Always ask why someone wants a report

• Reporting with a **Story** is Analytics

• Automate through business intelligence tools
  • **Reallocate** resources to analytics as you automate
  • Be sure to focus on usable reports
  • Report usage tracking is critical to knowing where value is being extracted
Start With The Analytical Talent In Your Company

• Who is the go to person the decisions makers
  • Find them, make them your champions

• Talent in your company already knows your business

• Over time, you will need to develop, migrate, or terminate talent
  • Talent decisions are the most difficult

• Hire strategically where you have gaps
  • First you need to know your gaps. Data Science will likely be one
  • Hire ONE data scientist at the start-likely from the outside
Data Science in the Beginning

- Start with the simple and fast
  - Regression works well in many cases
  - Many people can manage a regression
- Make sure you can explain the reasons for the predictions
  - Many techniques are more black box-avoid them at the start, build trust
- Ensure a control group is established
  - A control group enables rock solid ROI calculations
  - An enterprise control group is fantastic, if possible
  - You’ll get pressure to eliminate or not establish a control group
- Monitor your models

Image Credit: NASA/Reid Wiseman
**Systems Will Take Time to Improve**

*However, you must have quick wins*

- Data is all over your company
  - A search/inventory is the first objective
  - A future state **architecture** is the second
    - If you don’t have an architect-hire one. An architect has a different mindset than most
    - Make sure your architecture includes an analytical sandbox
- Initially buy top end **desktop or laptops**
  - Large drives, many processors
  - Your team probably has better computers at home. Don’t let that happen!
- Think of your 1st year baselining
  - Current state and future state vision with multigenerational plan to get there
  - Don’t just build a Hadoop cluster and dump data...you’ll **replicate** your swamp
Fund Analytics Through Business Growth

• Fund your team with “simple” wins
• Funding availability = business growth
• Start with CEO Strategy-It helps!
  • CEOs and Strategies change
  • Analytics must demonstrate ROI
    • Set up test and control groups
    • Spend time looking back and showing improvement
    • Always give the business credit for improving
Creating the Analytical Competitor

1. Speak the Language
2. Get Started!
3. Fund Through Growth