Enabling Enterprise Analytics: Part 2

Drilling for Responsive, Relevant, Rigorous, and Reusable Analytics

COL Matthew Benigni, PhD
Chief Data Scientist
Global Analytics Platform

02 January 2021

In 2014 Jason Toonders, the entrepreneur behind Yonego (https://www.yonego.com/nl/#gref), published "Data is the New Oil" (https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/) in WIRED magazine. He argued,

> *"Data in the 21st Century is like Oil in the 18th Century: an immensely, untapped valuable asset. Like oil, for those who see Data's fundamental value and learn to extract and use it there will be huge rewards."(Jason Toonders 2014).*

However, learning to extract information from data using advanced methods proves much harder in practice than in the classroom or in online tutorials.

In our last post we built upon Rogati's premise that production AI/ML requires organizations to progressively build upon a foundation of accessible, then usable data for experimentation by highlighting the need to extract value throughout each phase of data science hierarchy of needs to develop the organizational digital literacy needed to drive digital transformation. We described this in our last post as a "readiness mindset" where we use today's problems to progressively grow organizational data literacy.

In this post we will use a drilling metaphor to explain the relationships between motivating change, growing talent, and resourcing to maximize return on investment. We will also define aim points that help frame accountability metrics for organizational growth.

Drilling For Depth

The road map toward an AI-enabled force must pass through each phase of Rogati's data science hierarchy of needs. In the left panel of Figure 1, we invert Rogati's graphic, and in the left panel and illustrate a drilling metaphor in the right panel. If we compare extracting mission value from data to extracting oil from a well, we let depth be analogous to the level of methodological complexity our command can employ to deliver mission impact from data. In our metaphor we must "drill" through the phases of making our data accessible and usable before resourcing AI and deep learning into production data pipelines.

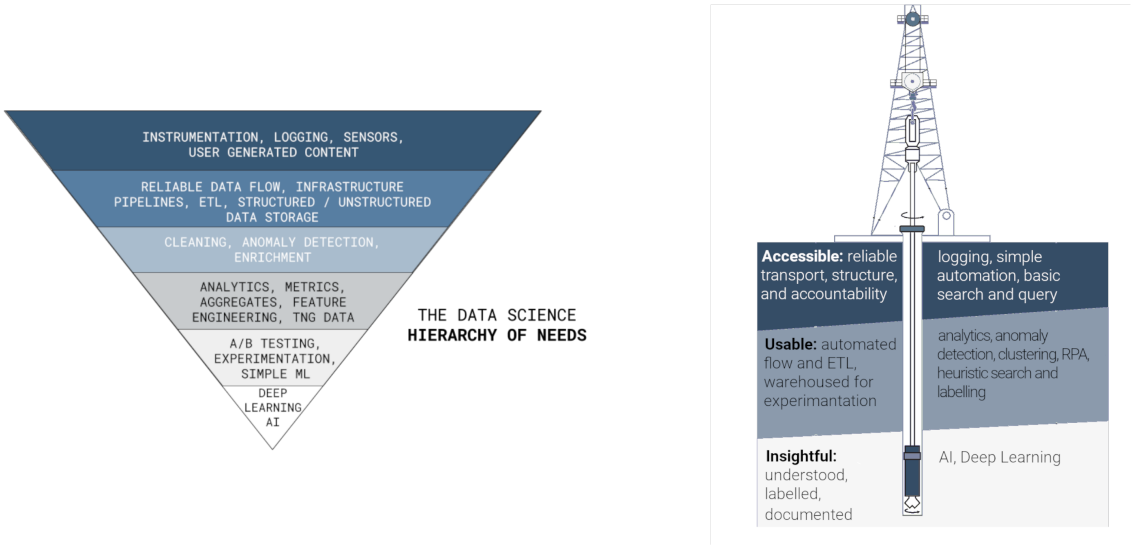

Figure 1: Drilling for Digital Literacy

The Physics of Operationalizing Data

Depth: *the level of methodological complexity the organization can employ to deliver mission impact from data*

Rotation: *an iteration of an analytic pipeline development cycle that produces novel insight and end-user feedback.*

Torque: *the ability to employ new methods and technologies for mission impact*

Friction: *effort spent delivering mission value that does not result in increased "torque"; wasted energy*

Premise 1: Drilling Must Start With Rotation

If we are going to drill towards implementation of AI and ML in production we first need to generate "rotation." We define a rotation as an iteration of an analytic pipeline development cycle that produces novel insight and end-user feedback. In other words an attempt to extract information from your data that solves a real problem. Ultimately our technologists need functional understanding of our military and intelligence professionals' needs to effectively select and implement appropriate methodologies and technical solutions. Our military and intelligence professionals need to become smart consumers of analytics. This shared understanding requires practice. Today's information management and business problems will deliver near-term value with a convenient down stream benefit, "organizational data literacy."

Measures Of Effectiveness: Responsiveness And Relevance

Responsiveness *- the quality of producing analytic results within a time frame that is usable by the customer*

Relevance *- the quality or state of being closely connected to mission need*

Generating this "rotation" can often be frustrating for senior leaders within the Department of Defense. We have an unconscious bias towards procurement, that prevents us from generating the rotation, torque, and friction needed to move towards AI and ML. Organizations must identify tractable problems that deliver value within an accountable timeline. We focus on **responsiveness** and **relevance** to maximize shared understanding between our operational leaders and technologists. Each have different responsibilities that require accountability.

- Operational leaders must be responsible to employ technologists against **relevant** problems. Ultimately our operational leaders are employing a novel capability, technologists. If they do not deliver mission value, they need to re-evaluate and try again.

- Technologists must ensure they take on relevant problems that are tractable and can be solved in a time frame that meets mission need. They must be **responsive.** Technologists must transparently estimate time to task and anticipated outcomes and clearly articulate when these estimates prove false.

This focus on solving **relevant** problems in a **responsive** manner ensures that operational leaders and technologists produce the organizational digital literacy that drives transformation.

Premise 2: Train Talent For Torque

In early phases of our digital transformation we focused on projects that deliver value (relevance) in less than 15-60 days (responsive). However, we emphasized areas with high potential for mission impact as we grow with respect to operational understanding, methodological breadth, DevOps practices, and user-oriented design. We looked for "far-sighted quick wins." In our drilling illustration torque is analogous to our analytics team's ability to employ new methods and technologies for mission impact and is a function of the team's relative mastery of:

- Operational Understanding - This is the "science/art" of your business. CRISP-DM (https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining) refers to this as "business understanding" (Smart Vision Europe 2019). The data science community often calls this "domain expertise." Your analytics team produces more torque when you organize it to understand where value/opportunity lie in your data.
- Methodological Breadth - Data science is an dynamic, incredibly broad field. Most data scientists have depth within a relatively narrow set of methods. Deliberately growing the methodological breadth of your team enabled it to more effectively employ novel methods.
- DevOps Practices- AWS defines DevOps as "DevOps is the combination of cultural philosophies, practices, and tools that increases an organization's ability to deliver applications and services at high velocity: evolving and improving products at a faster pace than organizations using traditional software development and infrastructure management processes. This speed enables organizations to better serve their customers and compete more effectively in the market" (Amazon Web Services 2020). We have found adoption of formal DevOps practices like Scrum (https://www.scrum.org/resources/what-is-scrum) (Scrum.org 2020) and Object-Oriented Programming (https://towardsdatascience.com/a-data-scientist-should-know-at-least-this-much-python-oop-d63f37eaac4d#:~:text=A%20data%20scientist%20works%20with,%2C%20classify%20medical%20images%2C%20etc.&text=The) (Mitsa 2019) to have powerful impacts on our analytics team's ability to deliver mission value. Our analytics teams strive to emulate the DevOps practices of production application developers within our organization.
- User-centered design (UCD) or user-driven development (UDD) is a framework that deliberately user needs and usability goals play an essential role in product delivery. Our analytics also emulate production development teams' ability to deliver products tailored to user needs and designed for intuitive use; however, we have found these principles also extend to technical writing tasks.

Part 4 (/post/building_analytics_teams_4/) will discuss how we use Scrum and technical writing to deliberately emphasize each of the four domains defined above. Within the context of this post the important concept is that an organization's relative mastery of each of these domains will drive both the complexity (depth) and velocity of your analytics teams contributions to mission.

Measures Of Effectiveness: Rigor And Reusability

> ***Rigor** - able to employ methodological complexity and deliver reproducible results while effectively communicating uncertainty, risk, and limitations*
>
> ***Reusability** - A workflow which can be used at scale repeatedly and/or for other use cases with sound data and software engineering practices*

While developing accountability focused on responsiveness and relevance is essential to grow an analytics team, long-term velocity and performance requires focus on **rigor** and **reusability.** offers a means to hold data teams and the customer's they support accountable for results, near-sighted focus on deliverables alone will not produce organizational digital literacy (Robbio n.d.). These near-term projects must drive changes in business process and inform infrastructure investment. Leaders must also focus on increased **rigor** and **reusability** of data science workflows.

Rigor often refers to specific standards of excellence expected from colleagues within a given profession. However, responsible employment can be challenging to quantify due to the breadth of the field and the pace with which methods and use cases evolve. Leader's must demand teams effectively communicate uncertainty, risk, and the

limitations associated with specific methods. Enforcing high standards allows maximizes professional growth while solving your organization's most pressing problems.

All too often, data science teams treat **reusability** as an after thought due to poor understanding of software engineering practices. For this reason, reusability is often the hallmark of a mature data science team. Some organizations use internal publication platforms like AirBnB's Knowledge Repo (https://github.com/airbnb/knowledge-repo) and NSA's Notebook Gallery (https://github.com/nbgallery) (NSA Open Source 2020) to foster reusability. Additionally investment in a shared data science platforms like Databricks (https://databricks.com/), RStudio Teams (https://rstudio.com/products/team/), and Microsoft Azure Learning Studio (https://studio.azureml.net/) can also significantly improve the reusability of data science workflows (Science 2019).

Premise 3: Resource To Reduce Friction. It Is More Expensive Than You Realize.

Premise 1 and premise 2 reinforce one another. Our ability to produce torque increases with rotation. With wise project selection responsiveness and relevance grow our rigor and reusability. Improves rigor and reusability result in greater relevance and responsiveness. Moreover, our ability to construct this type of problem solving environment aids in recruiting and retention. We want problem junkies. For these reasons it is critical to identify and remove barriers to "rotation" and "torque"; we use **friction** to represent effort spent delivering mission value that does not result in increased better operational understanding, methodological breadth, DevOps practices, or user-oriented design. Friction generally falls into four categories:

- Data accessibility and usability
- Secure access to and implementation of data science tools
- Barriers to end-user delivery
- Barriers to knowledge sharing

We prioritize infrastructure investment to reduce friction. In Part 4 (link) we will develop the concept of the analytic pipeline and how infrastructure can enable data accessibility and usability, secure data science tooling, end-user delivery, and knowledge sharing.

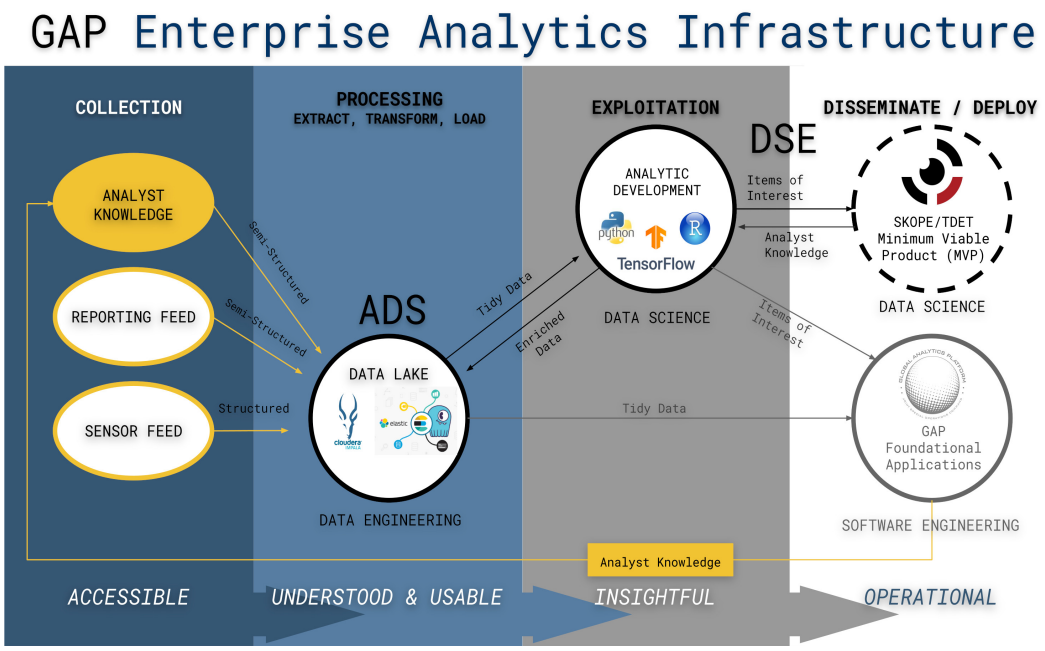

Figure 2: Analytic Pipelines and Platforms

Conclusion

In this post we used an oil drilling metaphor to explain the advantages of using today's mission problems to frame investment and build organizational digital literacy. We used "depth" to represent the level of methodological complexity an organization can employ to deliver mission impact from data. We then assert:

- **Rotation** represents iterative problem solving between an organizations analytics teams and operational customers. This process of iterative development with end-user feedback produces organizational digital literacy.
- **Torque** represents the analytics team's ability to employ new methods and technologies for mission impact and increases with rotation and command emphasis.
- **Friction** represents wasted energy, impedes iterative development (rotation), and reduced friction should be the purpose of infrastructure investment.

In our next post (/post/enabling_enterprise_analytics_3/) we will introduce the concept of an analytics pipeline and explain in detail how infrastructure enables mission impact and organizational change.

Works Cited

Amazon Web Services. 2020. "What Is DevOps? - Amazon Web Services (AWS)." ***Amazon Web Services, Inc.*** https://aws.amazon.com/devops/what-is-devops/ (https://aws.amazon.com/devops/what-is-devops/).

Jason Toonders. 2014. "Data Is the New Oil of the Digital Economy WIRED." https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/ (https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/).

Mitsa, Theophano. 2019. "A Data Scientist Should Know at Least This Much Python OOP." ***Medium***. https://towardsdatascience.com/a-data-scientist-should-know-at-least-this-much-python-oop-d63f37eaac4d (https://towardsdatascience.com/a-data-scientist-should-know-at-least-this-much-python-oop-d63f37eaac4d).

NSA Open Source. 2020. "Home Open Source @ NSA." https://code.nsa.gov/ (https://code.nsa.gov/).

Robbio, Alex. n.d. "Council Post: Is Robotic Process Automation Really A Driver for Digital Transformation?" ***Forbes***. Accessed August 14, 2020. https://www.forbes.com/sites/forbestechcouncil/2019/03/01/is-robotic-process-automation-really-a-driver-for-digital-transformation/ (https://www.forbes.com/sites/forbestechcouncil/2019/03/01/is-robotic-process-automation-really-a-driver-for-digital-transformation/).

Science, ODSC-Open Data. 2019. "10 Best Data Science Platforms." ***Medium***. https://medium.com/@ODSC/10-best-data-science-platforms-f59b41fd1958 (https://medium.com/@ODSC/10-best-data-science-platforms-f59b41fd1958).

Scrum.org. 2020. "What Is Scrum?" ***Scrum.org***. https://www.scrum.org/resources/what-is-scrum (https://www.scrum.org/resources/what-is-scrum).

Smart Vision Europe. 2019. "What Is the CRISP-DM Methodology?" ***Smart Vision - Europe***. https://www.sv-europe.com/crisp-dm-methodology/ (https://www.sv-europe.com/crisp-dm-methodology/).