*where **DATA** means **BUSINESS***

upside.tdwi.org

| Data-Driven Biz | Tech Talk | Intersection Biz-IT | Visualize It | Subscribe |

What's Holding Data Scientists Back?

By James E. Powell

August 25, 2016

Data scientists are the darling of tech reporting this year, but not every enterprise that hires them knows how to support them in their work. Upside recently spoke to **Kon Leong** about the role of data scientists. Leong is the CEO and founder of ZL Technologies -- a supplier of content archiving software for large environments.

Leong described how an enterprise can enable data scientists' best work by providing quality data management and encouraging collaboration.

Upside: How would you define the role of the data scientist?

Kon Leong: People tend to think of data scientists as number-crunchers, but this generalization vastly understates the true value of a skilled data scientist. It's n

---

SIGN UP — Get the Latest News

Latest Articles

- Alpha Centauri, Here We Come

- Data Digest: Analytics for Data Storage, Structure for Data Scientists, Build Scalable Applications

- Is Collaboration the Critical Success Factor for Data Science?

- Data Stories: The Impact of Global War

What's Holding Data Scientists Back?

SHARE WITH:

understates the true value of a skilled data scientist. It's not just about running statistical analysis on data; it's about knowing -- or figuring out -- the right questions to ask.

A curiosity engine is what gives data science its horsepower. Modern computing allows us to run incredibly complex calculations on incredibly large and diverse sets of data, but without the human capability to discern the right questions to ask, those computing capabilities are nearly useless.

In essence, the ideal data scientist is closer to a detective than a mathematician. The work is about collecting and processing data, discarding false leads, establishing multiple lines of inquiry, and piecing together the most plausible narrative.

What is holding data scientists back from executing the role that they were hired for?

We often see data scientists spending an inordinate proportion of their time simply trying to manage data: trying to make the data that's available into data that is actually useful and accurate. Data scientists need cleansed data to extract insight, and all too often the data hasn't been cleansed or managed before they arrive.

This is especially problematic in analysis of unstructured data: the messy content of email, documents, media, and other communications. Despite the need to manage this content for legal, compliance, and records management -- or perhaps because of it -- unstructured data is often scattered across multiple disjointed data "silos" throughout the enterprise.

The problem today is that in the big data era there aren't really important or unimportant types of content.

Q&A with Jill Dyché

Find out what's keeping teams up at night and get great advice on how to face common problems when it comes to analytic and data programs. From head-scratchers about analytics and data management to organizational issues and culture, we are talking about it all with Q&A with Jill Dyche.

View Recent Article

Submit Your Questions to Jill

What's Holding Data Scientists Back?

SHARE WITH:

In theory, these isolated repositories allow specific functions to be performed on certain subsets of data. For example, an enterprise content management platform might allow granular life cycle policies to be applied to certain "important" business records.

The problem today is that in the big data era there aren't really important or unimportant types of content. Data is simply data in the eyes of the law, and the organization is responsible for it. Isolated silos of content are preventing control and hindering data scientists.

Someone needs to be in charge of pooling and managing data so that it can be a resource rather than a burden. The question, then, is who should lead the way.

Who (or what) in the organization should be handling data management?

Data management isn't a specialized function requiring just a single specialist to maintain it. It's an enterprisewide framework that structures and supports the entire weight of the organization's intellectual capital.

A brick-and-mortar office space aims to maintain an organized environment where employees have all the tools they need to complete their jobs. Similarly, a comprehensive information governance framework provides easy access to all the human-created content that individuals need to conduct their communications and projects.

That said, *someone* needs to be in charge or else the initiative is likely to proceed at a glacial (or nonexistent) pace.

Many organizations in the early phases of a data management initiative will elect a committee to head the project, which is a step in the right direction. Often this team includes representatives from business units that work most frequently with human business content:

records management, in-house counsel, compliance, and
IT.

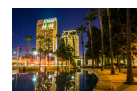The problem with this committee approach is that there is
often no "captain" at the helm; all committee members are
equally responsible, and they still maintain their full day-to-
day workloads. With that structure, it's very difficult to move
forward.

Nearly everyone in the organization depends on data
management in one way or another. That's why it is
absolutely essential to involve interdisciplinary stakeholders
throughout the process. Having a dedicated full-time
person in charge ensures that progress is made, that
stakeholder disagreements don't end up in stalemates, and
that all business needs are considered in the
enterprisewide data management infrastructure.

How can the C-suite better support data scientists in their work?

Stereotypes have uncanny persistence. We tend to think of
scientists -- in the "data science" context or otherwise -- as
lone masterminds rather than inseparable team players.
This is a dangerous misconception for the enterprise.

Direct lines of communication are critical, yet data scientists
often lack the support and collaboration that they need for
success. Data scientists are problem-solvers, not magicians,
and they need interaction with other stakeholders to
understand business needs, risks, and objectives.

Without regular exposure to a high-level perspective, they
will languish in the trenches with the data, missing the big
picture.

About the Author

James E. Powell is the editorial director of TDWI,
including the *Business Intelligence Journal* and *Upside*
newsletter.

**What's Holding Data
Scientists Back?**

SHARE WITH:

jpowell@tdwi.org

Upside Quick Links TDWI Quick Links

About Upside

Subscribe to

Newsletter

Become a

Contributor

Upcoming Events

Advertise

Contact Us

About TDWI

Upcoming Events

TDWI Research

TDWI Membership

Marketing

Opportunities

Onsite Education

Contact Us

Subscribe to Upside

Subscribe

What's Holding Data Scientists Back?

SHARE WITH: