EDM Webinar

Becoming a Data-Driven Organization – Transform or Perish

Live Date: June 6, 2023

Featuring:

Julia Bardmesser, CEO, Data4Real LLC
Eliud Polanco, President, Fluree
Peter Serenita, Chairman, EDM Council
Mike Meriton, Moderator, Co-Founder, EDM Council

Recording: View webinar

Presentation: View slide deck

EDM Council Homepage: edmcouncil.org

ADDITIONAL LINKS:

Download the "Becoming Data-Driven" white papers

Learn more about EDM Council trainings

Engage in EDM Council's upcoming DCAM Trainings

Engage in EDM Council's upcoming CDMC Trainings

WEBINAR Q&A:

Thank you to the panelists for providing the below answers to all questions posed during the live webinar. For more information or additional questions, contact us here.

With respect to cultural barriers, how do you see data literacy as a barrier, and what are ways to address this? What does a strong data literacy program look like?

There are several levels of data literacy. The one that's fundamental to the transition to the data driven organization is strategic data literacy - i.e. understanding by the business leaders how data capabilities enable achieving the business goals of the organization. That understanding is still very much missing in many data informed and data enabled organizations, however, without it, the transition to the data driven stage is not possible.

Strategic Data Literacy starts with making the business case to the business. So first they need to understand the benefits that will be derived from being data-driven and once they buy into that, then they need to understand what they need to do and what their responsibility is to make it happen. Trying to sell data responsibility without first getting buy-in to the business outcome has been a problem of the past and is why sometimes data programs get a bad rap.

The end result of an effective data literacy program is that data becomes very much the responsibility of the front line business folks, as opposed to being considered an Engineering or Back Office responsibility. This means that business people in the front are active participants in describing their data in terms they can understand and contributing to the business glossary. It also means that they participate in identifying when data is wrong and having a more active role in the remediation process. (It doesn't mean that they themselves are implementing code changes, but they have accountability for making sure good data is made available for their area by contributing if something is wrong what the right answer is).

Can you elaborate on the concept of "every consumer is also a producer"? In our case, not every consumer creates new assets to be "consumed" by other business areas.

In a data driven organization, every Consumer is a current or potential Producer. The idea is that any piece of data that is produced by any business function could potentially be of value in the future to any other business line (whether they knew it or not). Therefore it is everyone's responsibility to make sure that any data they create could potentially be consumed by someone else while the consumption use cases haven't been explicitly defined yet. This means that every piece of data needs to be described using terms people find and understand, and that there are clear policies defined on who cannot see data and when they cannot see the data whenever it is being created. (Otherwise the default position is to share it). For example, Finance is

traditionally considered primarily a Consumer of data from across the enterprise and rarely as a Producer. However, they do produce profits/loss, risk weighted averages, etc. that would be very useful at the micro (e.g. customer level) when making lending decisions. In a Data-Driven organization, Finance would therefore be considered both a Consumer and Producer.

Could you please explain with an example, including use case, technology details etc. how "AI powered Semantics" helps an organization become data driven?

One of the foundational principles of being a Data Driven organization is that everyone knows what the data means, so that they can then find and access it. In most Data Enabled or Informed organizations, business people manage glossaries and business context in one place (e.g., the Data DIctionary), and technical people manage the technical meta-data. To transition to Data Driven, we need to compress the business and technical together, so that the technical definition of a piece of data (e.g., its schema, table or column name) is coupled together with its business meaning.

But for this to work, someone needs to do the work to stitch them together. For example, if Finance defines something as "Party Name", and Risk defines it as "Counterparty Name" and the revenue producing business line defines it as "Customer Name", some process needs to link specific table or data set attributes to those terms and define that in certain contexts they are synonymous. That is very difficult to do by hand given that there could be multiple business vocabularies, thousands of tables and millions of rows.

The only way to scale this is through automation and intelligent systems. For example, at Fluree we have been using AI/ML models to scan and discover different business glossary contexts out of data by introspecting its contents and meta-data, and then use human expert feedback to detect synonyms and mappings between tables and their corresponding business meta-data attributes. We then save the discovered relationships in a Collaborative Database built on top of a Knowledge Graph, which makes the data instantly accessible to others based on rich security policies. That is but one example of using AI Semantics to enable transformation to becoming Data Driven.

How would you recommend that a very large company with thousands of systems and hundreds of data repositories move to this data-driven model?

Just like the classic answer to the question: "How do you eat the elephant?" "One bite at a time." It is important to note that a full transformation of legacy systems is a multi-year journey. However, we often recommend starting with use cases where we know that data needs to be shared across multiple parties and instantiate a new project/investment to create shareable data assets in a Collaborative Data Platform. For example, most organizations have a customer transformation program or a Customer 360 program where they need to share customer, account, product and transaction data securely to many parties, such as to Sales, Marketing, Product Development, Risk, Finance, Compliance, etc. We would start the new initiative there and select one or two domains as the starting point. If there is an operational Master Data Management System for entities like Products or Customers, we believe that those are good places to start to transition to these new Collaborative Data Platform frameworks. The important thing is to use these opportunities to modify the data flow so that instead of having a one direction process where systems update their own records, which then get fed into an MDM system, which then get looked up by other systems as a reference copy. Instead, the apps consuming mastered data from a Collaborative Database (whether operational or analytic in nature) are natively writing and reading into them as part of their application design. Over time, as more and more operational and analytic systems consume directly from the Collaborative Database, you can then start decommissioning legacy warehouses.

Do you think the average manager has enough knowledge of analytics to understand the importance of becoming data-driven? And thus actively participate in the required cultural change?

From our experience, to date the successful culture changes have been more top down where the senior business Executive Committee Level has either (a) identified data as a strategic asset with intangible book value, or (b) has identified the lack of ability to fully leverage its data as an existential threat. From there, FOMO usually takes over and this drives the cultural change required. But the initial drivers are usually not bottoms up or organic; they tend to be driven by a strategic burning platform to act.

Also, we think that Data Driven is not just about analytics; it covers everything from day to day operations to financial reporting to analytics. What's important for an average business middle

manager to understand is how his specific goals/objectives tie to data and data capabilities, including analytics.

Please provide some technology / tool examples of how zero-copy integration can be achieved? Does it refer to technologies like Data Virtualization?

There are two underlying principles to Zero Copy Integration: (1) the data needs to be stored in a secure data network as opposed to within a single host system. This means that the information is distributed by nature and that any member of the network can access the data without needing to physically move it to process it as long as they have the permissions encoded by contract or policy. (2) the data needs to be describable and have the ability to be queried using multiple semantic contexts, such as retrieving data from both Oracle Financials or Salesforce by using the terminology from either system or from a third, business friendly vocabulary. Data virtualization technologies was the first step towards this vision, but Zero Copy Integration mixes data sharing protocols and networking with Semantic Knowledge Graphs to achieve the vision. At Fluree, we've started to operationalize some of these technical concepts into real world solutions and would love to follow-up and discuss how they would work in more practical terms.

We are seeing efforts at digital transformation or digitalization of ops and/or workflow are taking too long. The 'product' - a viz app, digital twin, etc. - takes many months to launch, by which time markets have changed. How can leaders accelerate the time to launch data products?

The important thing is to take advantage of robust new technical capabilities that have emerged to deal with the volume, scale and sheer complexity of the transformation required. These include:

- **Automated data discovery:** being able to quickly scan, profile, and classify data against many simultaneous ontologies or data dictionaries
- **Automated content harmonization:** being able to discover duplication of data about entities and recommend how to reconcile the differences
- **Automated data observability and quality management:** being able to detect when changes to data are creating data quality drift and recommending how to remediate emerging data quality errors
- **Automated data lineage discovery:** being able to detect based on data pattern changes across multiple systems potential relationships that can be used to trace back the provenance of a data set

Given how fast data is being created, it's almost impossible to keep up with typical manual approaches. At Fluree, we're trying to pioneer technical standards and solutions around most of the concepts above. We'd love to collaborate with you as we carry some of these standards forward.

Do you take in consideration heavily governed businesses by the government like CMS?
Regulators tend to push towards adoption of frameworks and standards that make data common, correct and accessible, even when data is sensitive. In fact, one of the largest transitions we have seen to full Data Driven culture is in the US Department of Defense, whose CIO has authored a robust and strategic data sharing policy as critical to US national security, but in line with data privacy principles. Given how sensitive the data is, and how complex data sharing contracts, rules and enforceable policies need to be, the sophistication required to share data at this level goes beyond traditional data management systems. In fact, the US Government has led the innovations on development of new technical standards and frameworks to make secure Collaborative Databases viable. At Fluree, we're a member of the Working Group to commercialize some of these innovations. We recommend that everyone read the US DoD policy statement on Data Sharing as it is a good example of what top-down Data Driven culture transformation look like:

https://dodcio.defense.gov/Portals/0/Documents/InfoSharingStrategy.pdf