



White Paper

# **Analytics Collaboration: A Primer**

## Introduction

This paper examines the state of collaborating around data and analytics in organizations today. It also identifies and describes issues with current practices and describes how they should evolve in order to achieve an optimal data-driven culture within an organization.

As you read this paper, the most important thing to keep in mind is that analytics collaboration is a still nascent process where best practices and methods are still emerging. As organizations are evolving their analytics for deeper insights that answer increasingly complex questions, they are recognizing that collaboration can help them get there faster and with greater accuracy.

In order to discuss analytics collaboration, we must first define what it is...but also what it is not. Afterwards, we will discuss the different ways it manifests itself, the benefits that can accrue to organizations practicing it and, finally, the different methods of collaboration.

## Why the need for analytics collaboration?

While we just stated that the area of analytics collaboration is young, a good percentage of organizations claim to collaborate around analytics today. In a recent survey, data and business analysts were asked to describe how their team collaborates on analytics, with over 60% claiming to collaborate on most or all of their projects as can be seen in Figure 1 below.

### Q. Please describe how your team collaborates on analytics

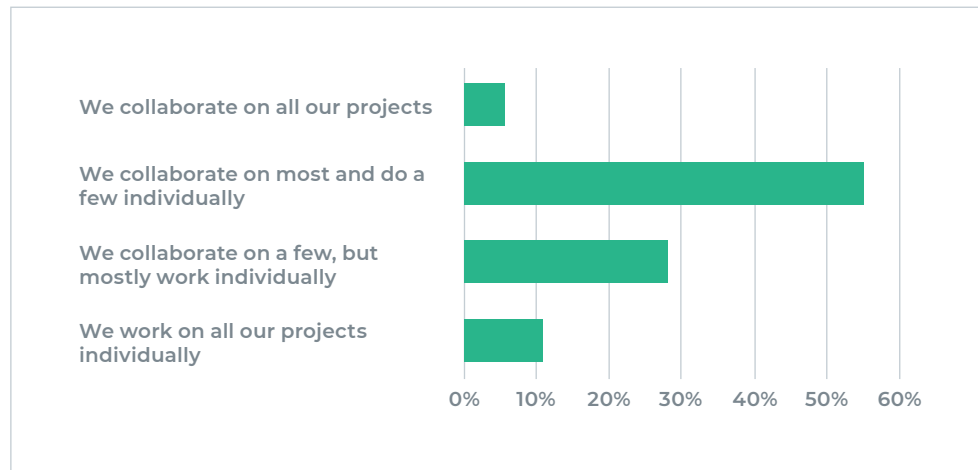


Figure 1: Analytics collaboration survey

The core need for analytics collaboration is to increase the analytics team's knowledge around data. A greater understanding on the analytics assets at their disposal facilitates greater trust, brings new ideas on how to use the analytics to derive new insights, and generally produces faster insights with greater detail and accuracy.

## What exactly is analytics collaboration?

The word “collaboration” itself has multiple meanings. In the most used one, it refers to the action of working with someone to produce or create something. This accurately captures what we mean when we talk of analytics collaboration; namely, the act of working with people in an organization to produce or create something of value with the organization’s data and analytics assets as the foundation.

We also note that we are focusing this paper on internal analytics collaboration among team members within the same organization. There is also collaboration with external business partners, clients and suppliers – which is also a worthwhile goal – but carries a different set of requirements and challenges which we do not tackle in this paper.

The very term “analytics collaboration” is itself in its infancy, with the industry not having fully coalesced around a canonical set of capabilities that define it. To that end, this paper examines the different methods by which analytic teams collaborate around data and proposes a critical set of capabilities required to fulfill analytics collaboration needs.

## Methods of analytics collaboration

Analytics teams, comprising of data analysts, business analysts, data scientists and management, will collaborate in multiple ways, all in an effort to share analytics assets and knowledge about those assets:

- Collaborating around discovery, creation, sharing and use of analytics assets
- Sharing personal knowledge of analytics assets
- Recommendations on how to use the assets based on personal knowledge
- Automated recommendations on best use of assets based on how others use them
- Shared automated (AI-driven) knowledge about the assets, including intimate understanding of the data, its uses, connections and other characteristics gleaned from user activity and detailed analyses of the underlying data
- Shared derived insights from the assets – both personal and automated
- Discussions, chats and collaborative documentation of assets
- Social feeds, sharing and discoverability of asset characteristics and insights

We will expand on these collaboration methods in the following section. But before we get to that, we also need to define what is not part of analytics collaboration.

## What analytics collaboration is *not*

Today, analytics collaboration is a process, set of best practices, and series of supporting capabilities inside of software products to facilitate the collaboration. Given the rather amorphous state of the collaboration capabilities at the moment, some vendors might categorize features from their business intelligence platforms as being part of analytics collaboration. But we would beg to differ.

For example, the following products or features of larger platforms are not by themselves classified as analytics collaboration:

- Enterprise data catalogs
- Source code management and code repositories
- Business Process Modeling (BPM)
- Shared data visualizations
- Shared enterprise ETL / ELT workflows

Some of these products might have a feature or two to support collaboration in some form but by no means do they cover all the aforementioned methods of collaboration.

## Types of analytics collaboration

As seen in the “Methods of analytics collaboration” section above, there is an abundance of capabilities that define an end-to-end analytics collaboration process. These capabilities generally fall within two distinct types: implicit collaboration and explicit collaboration. In this section, we present the basic characteristics of each type, describe their benefits and outline how they would appear in a collaboration platform. There is no implied assumption as to which type is better; both are valuable and provide benefits, and these benefits are multiplied in a process that can seamlessly blend both together.

### Implicit collaboration types and associated benefits

We consider implicit analytics collaboration methods to be those that emerge automatically and organically through an organization. These methods are typically implemented in a software platform by leveraging the knowledge of all team members’ actions and combining it with machine learning around activities, workflows and underlying data sets. They are also surfaced through the paradigm of social media, utilize machine learning and recommendation engines, and are centered around sharing and taking advantage of the power of collective team activity throughout the platform.

The importance of implicit collaboration to delighting analytic teams and enhancing an organization’s knowledge around analytics assets cannot be overstated. Implicit analytics collaboration helps to surface potential insights hiding inside the assets, create seamless workflows and ad hoc collaboration between teams, and generate additional exploration and discussions.

#### **Social media/ “Netflix UX”-influenced modalities**

These are recommendations on how to best use the analytics assets, under which conditions and for which purpose, exposed through a social media-style user interface in a collaboration platform. Based on implicit collaboration performed ad-hoc by analytic team members through traditional ways of sharing information (such as speaking face to face, over the phone or email) these approaches assume a software platform smart enough to recommend the proper uses of the asset to its users based on a variety of analysis factors. These include both the usage-derived recommendations we explore in this section, and data-driven insights which are described later. Discoverability and easy sharing are major drivers, and optimizing analytic team collaboration is the largest benefit. Some hallmark methods are listed below.

- Clear descriptions of each asset, its purpose, provenance, connections, potential uses, ‘best for’ recommendations and other relevant facts around the asset

- Recommendations around which assets to use in analysis and exploration
- Recommendations for similar assets to the ones currently being used
- Surfacing of relevant assets used by other team members through a Netflix-style recommendation engine and interface
- Display of interesting facts about all assets in the organization through a 'Did you know...' style of interface
- Identification of trending assets, and exposure through the interface, for possible consideration by the team
- Highlighting of assets recently added to or created in the platform
- A generalized social media-style interface combining aspects of the above into a holistic "newsfeed" interface for analytics assets
- Seamless discoverability and frictionless sharing throughout the platform. This needs to be designed in from the beginning as a major characteristic of the platform. Ideally, every piece of insight and information, generated both automatically and through team member intervention, should be shareable by one click and discoverable at a glance

### **Data-driven approaches**

These are additional recommendations, informational points and other actions based on deep statistical, machine learning, AI and other analyses of the underlying asset in a collaboration platform. These types of approaches are not feasible with the traditional ad-hoc way of doing collaboration that most analytic teams perform today. Below follows a short list of useful approaches.

- Automatically generated insights based on the data in the asset. While this feature may overlap with some of the newer capabilities in the self-service business intelligence space, it is also highly important to drive collaboration through unearthing connections and other insights inside the data, and exposing them to users. This should take the form of both informational facts around assets and recommendations for further usage and analysis
- Recommendations for additional asset exploration and the combining of assets into an „uber asset“

- Detailed statistics on the asset, including provenance, counts, usage characteristics, connections, popularity and trends
- Automatic alerts for significant additions and changes to the asset

## Explicit collaboration types and associated benefits

We consider **explicit** analytics collaboration methods to be those that require an explicit team member action for initiation or completion. These methods typically revolve around deliberate, team-driven requests, delivered through a variety of shared communication channels, which should leverage features of a collaboration platform to enable cooperative asset discovery, annotation and description. These methods are typically implemented through a combination of interfaces, including chat-style experiences, like/dislike (or “thumbs-up”/“thumbs-down”) feedback and a variety of documentation mechanisms. The overall experience is centered around explicit interactions between team members to better understand, document, curate and collaborate around the data. While manual, ad-hoc actions around explicit collaboration is definitely possible (but certainly cumbersome), the benefits of a software platform that codifies, centralizes and exposes these methods for wide use by analytic team members cannot be overstated.

Explicit collaboration is important to closing the loop between team members and the implicit recommendations described in the previous section. Through this type of collaboration, the team cooperates to deepen its understanding of analytics assets, above and beyond what the software platform used for collaboration has automatically generated. These processes also result in teams collaboratively discovering and documenting additional details around asset usage, provenance and purpose. Through this continuously enhanced feedback loop the team generates a permanent record that helps current and future members of the organization navigate and understand its analytics assets.

The main methods of explicit analytics collaboration include:

### Conversation and notes

Through the use of structured conversations, lightweight notetaking and annotation capabilities, a analytics collaboration platform should facilitate frictionless collaboration and decision-making. These notes and conversations can also be further analyzed through a variety of AI techniques. Some techniques and approaches in this area are listed below.

- Annotations and ‘case notes’ on a per project, product, asset, team member or other level, in order to document in place individuals’ particular insights derived from the asset



- Chat interfaces around assets, ideally within the collaboration platform, or captured through the organization's collaborative chat software (e.g. Slack or Teams). Machine learning and sentiment analysis can also be brought to bear on these conversation to further expose insights to the organization
- Structured email dialog around data, ideally within the platform, or at the very least captured though the organization's email system through tags or plugins. Machine learning and sentiment analysis can also be utilized to further derive knowledge from the organization's email exchanges

### **Team-driven documentation**

Through the use of shareable notebooks, wikis and other techniques, a canonical set of documentation around data can emerge through the collective understanding of the analytic team members. These will of course need to be shareable and discoverable and also serve as main input points in the cooperative asset inventory and discovery approaches described in the next section. Approaches for team-driven documentation include those in the list below.

- Shareable notebook-style data stories, centered around exploration of data, testing out of hypotheses and ideas, bringing together disparate assets, and presenting quick conclusions and interesting facts about the underlying data analysis. These assets serve as the basis for documenting exploration and its associated assumptions and insights, can be augmented over time by analytic team members and serve as basis for further development of a holistic data story
- Wikis, collaboratively edited documents that distill knowledge around data and can serve as the canonical source of organizational knowledge around a particular set of assets
- Light workflows, centered around connections between data assets, annotated by the team that serve as a quick way to discover non-obvious ways to combine data assets

### **Cooperative data management and discovery**

Building on documentation that the team has already created, as well as leveraging automatic and user-generated insights, analytic teams can work in a collaborative manner to analyze, prep and curate the organization's data. In addition, the collaboration platform's recommendation engine can be trained to further enrich and inform the analysis results. Typical methods of cooperative data management and discovery are enumerated below.

- Cooperative discovery and analysis of assets, utilizing common communication mechanisms to perform in-place data analyses, identify gaps and issues with the data and jointly improve the quality of results throughout the organization

- Asset modeling collaboration, either as a subset of full-blown data analysis or as its own standalone task. Teams will particularly benefit from shareable and documentable combining operations. A platform's role as either as a mean to combine assets with collaboration capabilities or as a hub to connect to specialized tools, layering on analytics collaboration capabilities seamlessly is important in advancing the team's knowledge
- Cooperative data modeling, in parallel and as a result of the above-mentioned discovery and analysis. Through the certification of data assets and data stories, and by using the shared documentation and communication processes already described, the team members should be able to certify data assets, analytic results, workflows, annotations, insights and all other automatically and user-discovered information about the assets. A byproduct of such cooperative curation is an ad hoc, highly accurate asset marketplace for the organization
- Training of the collaboration platform's recommendation engine by the use of a social media-influenced like/dislike mechanism for automated and user-generated insights, allowing recommendations, insights and discoveries of the assets to rise and fall in relevance and popularity as the team collaborates more closely over time

## The benefits of analytics collaboration

The descriptions in the previous sections of analytics collaboration processes and methods lead us next to discuss how they add value and cause benefits to accrue to an organization. We believe that there are six distinct axes around which benefits can be realized and for which the need for a analytics collaboration process becomes evident for an organization. These can be seen in Figure 1 and are described in detail further below.



Figure 2: The analytics collaboration axes

### Axis 1: Knowledge sharing

If we could pick only one reason to invest in a analytics collaboration process, it would be for its ability to enable broad knowledge sharing across an organization. Currently, most collaboration is done in an ad-hoc manner, often through endless email back and forth, resulting in knowledge output that is siloed, ineffective, incomplete and not easily shareable or discoverable. Analytics collaboration has the potential to significantly improve on this experience, removing barriers to entry and most of the friction involved in working with others and sharing discoveries and data insights.

### Axis 2: Deeper data understanding

By enabling more people to jointly work together on analytics and by centralizing the insights and information uncovered, the organization becomes wiser over time about its analytics assets and their associated characteristics, usage, connections and popularity. In addition, as the understanding around analytics deepens, new uses for the analytics emerge and can lead to opportunities that otherwise would not have been easy to predict.

### Axis 3: Increased Trust in Data and Analytics

Increased knowledge and transparency around analytics builds greater trust in how to use the assets and the information they produce. This benefit shows up with the analytics producers – analysts and data scientists – enabling them to produce results faster and with greater detail and accuracy. These benefits also bubble up to management – who can create data-driven strategies from the analytics with confidence, and execute at an agile pace.

## Axis 4: Enhanced curation

As the organization deepens its knowledge of the analytics assets, it is presented with a big opportunity to organize, tag, document and thoroughly curate its formerly uncategorized assets, as a certified set of canonical analytics assets. In addition, through the sharing and easy discoverability that a analytics collaboration process enables, these curated assets become widely available as golden sources for myriad uses, displacing the current ad hoc, local, manually intensive and only partially accurate methods of local data extracts and marts.

## Axis 5: Cataloging and Inventorying acceleration

As many organizations who have invested in data catalog software in the past few years can attest, getting an accurate, continuously maintained catalog of golden source data to be relevant and used is a mighty task. Analytics collaboration methods significantly accelerate adoption through the collective power of team members' analyses, iterative and perpetually improving processes and the easy discoverability and sharing that is at the core of the methods described above. In addition, analytics collaboration covers any analytics asset, not just data, helping teams link any of these assets to results and their analysis at hand. Organizations can move away from manually-maintained, infrequently-updated and sparsely-used catalogs and leverage the power of analytics collaboration to always access the most accurate, best-maintained and meticulously-documented analytics assets.

## Axis 6: Enhanced data provenance tracking / auditability

Lineage, master data management, governance, traceability, provenance, auditing, as well as other, related technologies and their associated software platforms, are all in use today, across a variety of organizations, to mostly middling levels of success. An analytics collaboration process – by virtue of its design as a user-powered set of methods and tools that make it easy to discover, share and curate analytics assets – is in the best position to answer and document questions related to the facets mentioned above. In addition, by leveraging the process to enable a widely-used, centralized asset marketplace in the organization, significantly increased accuracy and usage can be achieved.

## Conclusion

Analytics collaboration is a complex, still-evolving category that, while serving a real need in an organization, has not yet reached maturity or gained critical mindshare in the industry. The situation is rapidly evolving, however, with the benefits becoming increasingly apparent to organizations of all stripes, the need for well-defined collaboration processes becoming better understood and a software ecosystem gradually emerging.

In the near-term, the industry will begin to mature and dominant vendor products will establish themselves. A seamless combination of implicit and explicit analytics collaboration capabilities will remain the holy grail of the market for some time.

Vendors that enable the processes described in this report, who invest in a frictionless user experience, and avoid simply rebranding their business intelligence products, will establish themselves as leaders in the marketplace. Similarly, customer organizations that fully buy into the ethos of end-to-end analytics collaboration will realize most of the benefits, enhance the quality of their data, build a genuine data-driven collaboration culture and gain significant insights into their business through a combination of robust processes and world class software platforms to enable them.