GIGAOM RESEARCH

The Benefits and Challenges of Personal Analytics

David Loshin

February 12, 2015

This report is underwritten by FirstRain.

TABLE OF CONTENTS

Executive summary

Presenting business analytics attuned to individual needs—what we call "personal analytics"—is often difficult, but deployed properly it can deliver results that inform decision-making fully aligned with corporate strategy. Doing so requires conquering the challenges associated with adding context to content.

A clear distinction should be made between personalization and personal analytics. Personalization means crafting the presentation of material to meet individual expectations. Personal analytics, meanwhile, is the filtering and organizing content relevant to each individual's scenarios and then leveraging that person's preferred methods of accessing and absorbing information to streamline the business and decision-making processes.

The report will help CIOs, CTOs, line of business managers, and other IT decision makers understand the importance of blending data, content, and context, and will give them guidelines for assessing tools for personal analytics.

Key highlights in this report include:

- Flagship consumer apps have increased expectations and created a need and opportunity for enterprise software to deliver personalized experiences inside traditional applications and workflows.
- When all individuals in areas of expertise use the same search engines, search strings, and industry journals, they'll derive the same intelligence about what is happening in the business, industry, and market.
- Personalization, an assumed part of the technology user experience (UX), is key in solving this the problem of data relevance.
- The ability to transform numerous data sources into personalized intelligence feeds involves methodology that is likely to tax most IT shops.
- Engaging external vendors to augment internal business intelligence with personalized analytics is a way to decrease time-to-value.

The Challenge of Democratizing BI: Finding Relevant Information

Everyone in an organization can benefit from actionable knowledge when it's easily accessible and timely. Cutting through the big data analytics morass to see what's missing is the basis of democratizing business analytics. Organizations now face a new challenge: data relevance. Self-service business intelligence (BI), analytics and end-user data visualization tools have established the framework for delivering knowledge, but how does an individual in an organization absorb the business, industry, and market intelligence applicable to her or his specific background, role, and expertise?

An increase in content has magnified this conundrum. Although readers are bombarded with news and articles from multiple sources as well as information that is copy-pasted or slightly rewritten, persistent content remains at the top of the pile. Finding relevant content deeper in the stack is increasingly difficult, not least because it's hidden or disguised—when, for example, a company understates the importance of news such as management changes. The task of differentiating between information that needs attention and that can be ignored has increased most business people's workloads.

Personalization, an assumed part of the technology UX, is key to solving the problem of data relevance: People succeed in their business roles when they can apply their own knowledge and experiences. Organizations wanting to give their employees access to personalized content must understand what makes each individual unique and then apply that understanding to filtering and framing inputs.

Consumers—these same business individuals outside of work—are now accustomed to a variety of consumer apps that personalize the user experience, but most organizations still default to a cookie-cutter approach when delivering content to end-users. Within any area of expertise, each person uses the same search engine, the same search strings, and the same industry journals. It's little wonder, then, they ultimately only have the same intelligence about what is happening in the business, industry, and market.

The Value Proposition of Personal Analytics

Personal analytics extend beyond superficial or cosmetic rearrangement of information to focus on the individual: who that person is, what that person does, and how that person fits within the organization, and to what degree the organization depends on that person's decisions. The objective is to filter and organize the content relevant to each individual's scenarios and then leverage that person's preferred methods of accessing and absorbing information to streamline the business and decision-making

processes. For example, periodic sales reports are meaningful to a variety of roles within a business. But the "spin" accorded to the specific collection, organization, and presentation of sales reports will differ according to individual roles in the organization: The CEO may want to verify that sales accord with the forecasts by time and location; the VP of marketing may want to correlate periodic sales to marketing spend by medium and venue (e.g., local TV commercials vs. highway billboards); and a salesperson may look for gaps in exploiting predicted revenue opportunities.

More data enables more robust analyses and more precise reporting, but a naïve approach to absorbing and utilizing the flood of data can overwhelm a business analyst and obfuscate critical pieces of knowledge that are relevant to one individual and meaningless to others. For example, the full-scale sales report provided to the CEO may be too general for the V.P. of Marketing, but overwhelming for the regional sales representative.

Personal analytics has evolved out of a need to consume tremendous amounts of information in a very brief time. Analyzing business information and sharing it—or selling it—is no longer restricted to research professionals. Those who consume business information in a company have different roles, different needs, and different—though always decreasing—amounts of time for research and analysis.

Personal analytics provides a balanced "two-lensed" approach. The internal lens analyzes massive amounts of information and organizes that information. The external lens views relevant results based on business context, role, and the information consumer's immediate need. Combined, the two lenses provide a way to present actionable knowledge to specific classes and categories of users—organized based on their roles within the organization—in ways that help drive profitable business decisions and actions. As an example, one organization might scan through Twitter and Facebook posts for references to corporate brands and another customize the reporting of brand monitoring activities to individuals in relation to the actions they are empowered to take with respect to brand protection.

The Challenges of Adding Context to Content

Any approach for personal analytics must adapt to changes in the way individuals work within their business context, especially with the growing scale of data volumes. Organizing information that is customized in a way that makes sense for each business consumer requires imbuing content with context before maximizing the value of contextual filtering and presentation.

Optimizing massive amounts of data and crafting a presentation of actionable knowledge for each business consumer's needs presents many challenges. We'll separate those challenges into data analytics challenges—developing algorithms that will collect, analyze, group, filter, categorize and ultimately extract meaning as well as continuously retrain the machine and humanize the selection process—and personalization challenges—slicing and dicing this information based on personal needs and delivering it in a way that is most helpful depending on a person's state of mind (location, time of the day, device, etc.).

Data Analytics Challenges

- **Content variety**. Data sources are no longer strictly structured. Business people increasingly rely on a pool of data artifacts that blend traditionally structured data with numerous types of artifacts, such as transaction system databases as well as social media channels such as Twitter, Facebook, LinkedIn, blogs, wikis, and expert channels, each of which must be assessed for contextual relevance and integrated within the various information models.

- **Content quality**. The actionable bits of information that can be extracted from a data source like a database or a social media site may have different levels of relevance for different types of data consumers in different places in an organization. One example is data collected for reporting about product launches; for the senior executives, a rolled-up perspective of positive or negative sentiment may be sufficient, while the product manager may look for details regarding potential product flaws that can be rapidly remediated. A complication of ensuring the proper assignation of meaning for each data artifact within many different business contexts is assessing the quality of the sources for each of the various contexts in which they can make sense.

- **Content organization**. Organizing data inputs begins with a set of models for semantics and meaning, but business needs change over time, so the models must be adaptable with an ability to tag inputs, provide weightings in relation to taxonomic models, and match them based on inferred content. However, different levels of information density, sparseness, quality, and freshness (among other factors) influence the ability to organize data and require increasing sophistication.

- **Connectivity.** Any data source might have different levels of relevance within many different business contexts. For example, comments about a car's drivability may be more meaningful coming from an automobile enthusiast web blog than what can be scanned via Twitter. That suggests two challenges: the first involves connecting data artifacts to different business domains, while the second involves inferring dynamic linkages, relationships, and relevance beyond fixed taxonomic models. The latter challenge also means attempting to evolve an understanding of how data sets are used by different individuals and adapting the analytical models accordingly.

- **Making data useful not flat**. This refers to deriving information that was not presented in a standard way but rather discovered through understanding the structures and relationships. For example, a company may have a number of different lines of business, each with its own sets of products and services, and that company may compete with other companies in different markets. Among other bits of information, corporate transactions, mergers and acquisitions, news items, and information about new corporate hires can all contribute to a hierarchical reflection of a competitive intelligence profile. This is not only helpful for identifying when competing companies are entering those markets with new products or services, but also the methods by which the competitors will enter the market, potential launch locations, and even how corporate partnership arrangements are intended to promote the sales and marketing efforts.

Personalization Challenges

- **Filtering**. More important than filtering through large volumes of data assets taken from a variety of sources is that many different filters must be in place to discern different types of business value depending on who the consumers might be. For example, a sales representative might be informed as to deep knowledge about specific contacts within their customer and prospect base to help in the direct sales efforts. The same data sources can be filtered to provide sales and marketing executives with qualitative information about their top customers, help to identify imminent threats from potential competitors marketing to those customers, and inform strategies for continued expansion within vertical markets. After initial organization and filtering, the next step is feeding the selected artifacts into analytical models tailored to meet the needs of specific individuals, within specific business contexts.

- **Understanding the nuances**. What might be deemed a "skepticism filter" can be introduced to weight the believability and trustworthiness of particular data sources. Introducing these types of skepticism measurement is important for layering aspects of a human perspective and providing a

perception of managed mistrust to accompany automated analysis. These filters may differ depending on the data consumer, lending more credence based on how each role is expected to consume information.

- **Finding correlations in a dynamically changing business world**. Uncovering patterns of data correlations may indicate emerging trends. An example might be examining the correlation between web searches about flu symptoms and remedies and geographic locations over time can help in predicting the infection patterns for influenza.

Use cases

Personalized business analytics aims to deliver the right types of information to each individual playing a particular role. Some high-level use cases can demonstrate how filtering and differentiation by consumer context can simplify the exploitation of business analytics:

- **Customer intelligence for sales representatives**, which blends internally managed customer data with customer intelligence drawn from among a variety of external sources. This fused customer intelligence can be packaged for presentation to end-users, but can be further personalized for each sales representative to include information about similarities to existing accounts, identification of individuals within the sales representative's professional network, as well as the prospective customers' business issues and preferences.

- **Identification of new markets**, which accumulates data about market drivers, business conditions, and perspectives from key industry influencers to not just discover new market opportunities but to also help *time* the entry into those new markets. This information can be presented in one way to product managers seeking ways to enhance their existing offerings, in a different way to the marketing teams to motivate planning and creation of campaigns, and in yet a different way to sales account representatives.

- **Improve the revenue cycle** by providing insight as to the various times associated with the different phases of the "order to cash" process. Specific metrics, enhanced with additional customer insights associated with the different phases of that process, can be shared with different team members in ways that can shorten the overall process and improve cash flow. For example, customer predispositions to purchasing can be shared with the sales team, while information about credit-worthiness, compliance with purchase terms, and general risk can be accumulated and presented to the finance group. At the same time, supply chain information can be blended and filtered to the fulfillment and delivery teams.

Blending Data, Content, and Context into Actionable Knowledge

Most approaches to content filtering and organization focus on predefined concept hierarchies coupled with text analytics algorithms that meta-tag the data inputs. Yet attempting to stream, parse, interpret, and organize all the data coming from numerous different sources establishes a very broad scope of effort, especially when not all information carries the same relevance for different data consumers. A more effective approach would introduce the processes and technologies necessary to narrow the scope to focus on business-relevant information aligned across specific business domains using the end-users' characteristics and preferences as a guide, using the following three steps.

1. **Ingest a rich and broad variety of data sources.** Start with multiple information sources, such as corporate and personnel information easily accessible from company web sites using straightforward web crawling capabilities. Add media news outlets, syndicated news sources, blog sites, wikis, expert channels, press releases, as well as regulatory filings and mandatory reporting artifacts. Content streaming from social media sites such as Twitter and Facebook can also be included.

2. **Make sense of each of the data assets.** Ensure that the data's embedded characteristics can be parsed, identified, and extracted. Stream-processing technologies can be augmented with text analytics to scan streaming content and filter out meaningful content from "noise." Semantic analysis algorithms can peruse content to identify key entities that might be relevant to any of the downstream data users, including individual names, corporate names, industries, product types, business lines, , geographic locations, behavioral characteristics, and other conceptual topics.

3. **Tag content items.** Tags should be based on their embedded concepts in relation to their level of business-relevance. Take advantage of concept hierarchies, taxonomies that organize terms and phrases based on the concept hierarchies, and contextual interpretations that can inform the filtering and packaging downstream.

The process must go beyond organizing content according to *de facto* industry standards. Instead, the goal becomes making use of an organization's analytical capabilities for filtering, indexing, and presenting content to each business user. That suggests a need for information models that evolve in relation to how the business user works and succeeds.

To ensure proper personalization, it is not only the streamed content that needs to be processed and analyzed. Precise and accurate personalization depends on high quality "live" analytic models that link concepts in relation to a collection of contexts potentially associated with any individual user, including that user's experiences, preferences, educational background, employment background, "street smarts," as well as the one or more roles an individual might play within the organization.

To enable this within any organization, one would need tools that can accommodate all the technical aspects implied. That includes accessibility to a broad palette of data sources, the ability to ingest and process that data, organize the data within known conceptual contexts, and represent that information in a more structured manner that allows analytics applications to package that data in a way that can be presented to different business professionals in the ways that they can best take advantage of that information to execute specific tasks.

Assessing Products and Services for Personal Analytics

While some organizations may presume to develop some of these techniques in-house, developing a robust trustworthy capability may tax even the most sophisticated IT organizations. Rather, a preferred method is to select a vendor providing these capabilities. Although many BI and analytics tools providers are evolving these capabilities in their products, the reporting and analytics capabilities provided by modern business intelligence tools are more likely to address aspects of *presentation* personalization, not content personalization. In other words, they may not satisfy the need for assembling the information into the distinctive perspective for each individual within that individual's particular context.

Organizations will look to adapt their own preferences, scoring algorithms, analytical models, and data ecosystems by integrating externally developed techniques to provide more robust personal analytics. The first step here clearly identifying the business needs, such as the ones associated with our previously referenced use cases, such as customer intelligence, market filtering, or risk management. The information gathered establishes the baseline criteria for evaluating and comparing solutions to address an enterprise's needs for personal analytics. Soliciting the business needs centers on understanding the opportunities for personalization, and some of the steps include:

- Assessing the breadth of user expectations within the organization, associated with customer analytics, market landscapes, and marketing insights.
- Engage each of the users and ask about their background, expertise, and their perceptions of the gaps in information presentation to inform their daily activities.
- Observe how each individual works within her/his business environment and how he/she uses information to inform day-to-day decisions.

When evaluating tools for personal analytics, consider these capabilities:

- **Big data.** As data volumes from sources like Twitter or other social networking channels continue to grow, the potential for identifying data that may be relevant to multiple contexts and individuals increases. The demand for personalization may drive the desire to absorb greater data volumes to increase competitive advantage. Ensure that any candidate tool is able to handle the

potential rampant increase in data size. Look for tools that can monitor and absorb both static and dynamic (streaming) data sources simultaneously, and that can scale for massive data volumes.

- **Data variety.** Relevant content can be derived from many different sources, few of which conform to established structured data formats. Make sure that the tool can easily consume and interpret the content within both structured and unstructured data from both the open as well as the deep web

- **Semantic intelligence.** Interpreting the subtleties of meaning is critical and requires methods for content parsing, contextual entity identification, concept assignment, scoring, and organization within and across different business domains, industries, companies, business topics, as well as dimensions such as time and location.

- **Real time.** Depending on the business context, the window of opportunity may impose constraints on the time for processing and data availability. For example, a business strategy team may have a limited amount of time to engage an acquisition target and propose an acceptable offer before a competitor swoops in. With compressed real-time windows, value is added when using tools that can rapidly absorb data from a wide variety of data streams, interpret, filter, organize, and then present it to the business user within the acceptable time frame.

- **Consumer-orientation.** The tool should be able to manage a content-organization framework incorporating profiles customized for unique individuals. It should also be able to filter information in relation to the user/consumer model within each individual's personal business context.

- **Prioritization.** The analytics engine must be able to assign a level of importance to information, based on different information consumer profiles.

- **Adaptability.** The underlying profile models, as well as the presentation of information, must adapt as an individual's personal context evolves and changes. One example is when an individual within an organization changes roles—for example, a sales person being promoted to regional account manager. That individual's *personal* preferences for information may stay the same, but because the role has changed, the presentation must be adapted to fit that new role. Similarly, as different individual personalities are perceived to populate a particular type of role, the overall personalization characteristics may evolve in relation to the underlying individual personalities.

- **Feedback.** The tool should be able to blend user feedback and emerging trends derived from the source data back into the framework to enhance and improve the personal data organization models. For example, associating business metrics with individual actions can indicate whether some facets of the personalization filtering are more "productive" (that is, lead to more desirable results) than others by comparing how each individual reacted to tunable aspects of the presentation.

- **Self-organization.** Although the scope of predefined business concept hierarchies may be limited by prior experiences, data organization algorithms must be suitably self-aware, using approaches to determination of importance and relevance using self-organization techniques informed via the user feedback collection. For example, news media channels and aggregators indicate relevance and popularity of content in relation to the frequency, timeframes, and volume dedicated to specific news stories. The more frequently additional information is requested about a particular topic or article, the greater its relevance, while those concepts subject to diminishing interest can be deemphasized.

- **Proactive**. Self-organizing models can infer dynamic contexts and consequently data importance, that can be used to identify opportunities to alert business users to information they might have missed—and to which they must pay attention.

- **Workflow.** The tool should blend into the individual's workflow as well as with workflow tools used within the organization.

- **Content adaptability.** The tool should allow the quality, amount, and presentation /visualization to change based on the consumer conditions at any time during the day—access to device, amount of time, etc.

- **Sharing.** The tool should allow the team to share information, especially if there are similar roles filled by individuals with similar backgrounds. This can help "jump-start" the personalization for new users.

Key takeaways

- The availability of an increasing number of data sources increases the potential of accessibility to information adding additional business insight or enhancing customer intelligence.
- The results of absorbing and processing these data sources are just a preliminary step to adding business value; the final mile is actualized by presenting the information in the way that is optimized for consumption.
- Different individuals take advantage of actionable knowledge in different ways, but all are likely to embrace analytics when it is personalized based on a combination of experience, preferences, role, and context within one or more business environments.
- Tools that can blend many data sources must be able to ingest massive amounts of data from a broad palette of sources, have internal methods of semantic concept organization to identify relevant data and sort it into organized contexts, as well as customization methods to personalize the presentations based on user profiles.

An emerging number of vendors are developing services-based approaches to personalized analytics, and adopting these vendor offerings will speed time to value.

About David Loshin

David Loshin is recognized world wide as an information management industry thought leader. He has dedicated many years to popularizing best practices for business intelligence, data governance, performance computing, master data management, predictive analytics, and data quality.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.