Trendex: Telling Stories through Data Analysis and Presentation

Welmin R.D. Militante

EXECUTIVE SUMMARY

Data scientists want to believe that data have all the answers. However, the essential element to accompany extracts, data universes, queries, and quantitative syntheses is qualitative analysis: asking questions; creating directives from data; and telling its story. Sonoma County Human Services Department developed a tool wherein "monthly report(s) of current and trended workloads, outcomes, and staffing [levels]...inform [staff], partners, and the community about the breadth and depth of [their] services."[1] Over time, Trendex has developed into a vital tool that is used organization-wide for the purposes of improving operations based on measurable outcomes and making informed decisions that are data driven.

Santa Clara County stands to gain enormous benefits by considering emulation of the investment in and development of a centralized reporting tool to capture agency data, trends, and performance indices. While charts and tables are commensurate elements to communicating general information, data storytelling through comprehensive and effective analysis is an essential component in crossing the chasm between understanding important findings and the ultimate decisions and actions that complete the story.

1. Trendex Report – Purpose Statement

Welmín R.D. Militante, Application & Decision Support Manager, Santa Clara County Social Services Agency

Trendex: Telling Stories through Data Analysis and Presentation

Welmin R.D. Militante

Introduction:

Every Monday through Saturday, information systems departments across every CalWIN consortium county receive several gigabytes of information from the county's previous day transactions in CalWIN.[2] This translates to millions of data points for tracking an organization's performance. The challenge for data consumers is to translate this mountain of information into meaningful assistance for staff so that informed decisions are made—whether it's distributing new CalWIN cases to various Continuing Units to achieve parity and balance, or justifying a request for additional resources to address a growing backlog in the processing of new Medi-Cal applications as a result of the Affordable Care Act.

Compounding this dilemma, the volume of raw data that organizations have are becoming increasingly overwhelming. Of all the data in existence today, 90% were created in the last 2 years.[3] Inundated agencies stand the chance of losing sight of the difference between what is statistically significant and what is critical for decision-making. Using data successfully requires human translation and environmental context regardless of the audience toward which the information is geared. Without a human frame, blanket interpretation of raw data may lead to confusion, or worse, deleterious decisions. Accordingly, it will benefit management for Santa Clara County Social Services Agency to receive the type of data analysis that evaluates numerical outputs into an understanding of the organization, and as a product of the social/legislative/economical environment in which it operates. In essence, this act "humanizes" information by turning raw data into a story about a collective performance, and subsequently transforms "bits and bytes" into actionable directives.

Trendex

In 2007, Sonoma County Human Services Department (SCHSD) embarked upon a vision to develop a set of tools that provides decision-makers with holistic and robust information to drive thoughtful assessments and arrive at informed decisions. In order to achieve this goal, the Information Integration Division (IID) under the leadership of Jo Weber (former SCHSD Director) began to collect and organize data from various sources in order to measure performance and subsequently interpret programmatic outcomes. It was envisioned that this centralized approach to data collection and analysis provided a nexus upon which a consistent and uniform product (Trendex Reports) is produced. Prior to its implementation, SCHSD expended a great deal of resources to address the data analysis needs of its many departments, often with widely varied outcomes.[4] With Trendex, managers and subject matter experts from the services delivery divisions worked collaboratively with the IID staff to define information needs; identify data sources and how to collect information; determine required analysis;

2. Santa Clara County, for instance, receives between 5 to 10 gigabytes of data from the daily CIS download, exploding up to 100 gigabytes during Cost of Living Adjustment (COLA) runs.

3. IBM, "What is Big Data?" (http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html)

4. According to Jerry Dunn (SCHSD Director), there was a time when programs were designed based on intuitive knowledge backed with anecdotal information.

and validate the findings. The amassed data, some of which are manually collected, is entered into the Trendex database and spreadsheets, and the resulting reports, beginning with March 2009 release, provided a path to forming agency strategies, desired outcomes, and the robust processes that meet defined goals and objectives.

Report Presentation. The current iteration of Trendex is released in several sections. Every release defines for the user three elements that make up the intent of tool: trend (a general direction), index (a guide or reference that reveals or indicates), and purpose (a monthly report of current and trended workloads, outcomes, and staffing to inform stakeholders about the breadth and depth of services). Along with a "Table of Contents" to aide the user in accessing desired information quickly, the reports provide an overview of material changes to the report from the previous month's edition. This section ("What's New this Month?") provides a running tally of how the tool constantly evolves to meet the changing business needs, further exemplifying its status as a "living document." Finally, the "Highlights" section offers a quick glance at significant trends that transpired for tracked service categories along with color-coded directionals that accompany levels of change observed.

Interspersed among monthly counts of various data elements and trend information across 6, 12, and 24-month timeframes are summary graphs of key trends. These colorful plates call to the reader's attention the "story" that is happening in areas with significant organizational interest. The graphs provide a pictorial method to represent measurable elements across a period of time, informing users of past performance. That said, what truly makes Trendex an invaluable document is the detailed "Definitions" matrix; every data element is named, defined, source identified, and its collection method is documented. This built-in "data forensics trail" allows for ready auditing and cross-referencing. Users are not as likely to scratch their heads in an attempt to interpret information or guess as to how data are derived.

The Main HSD Trendex is produced monthly, covering agency-wide information along with subsections covering the Office of the Director; Adult and Aging; Economic Assistance; Employment and Training; Family, Youth and Children; and Fiscal-Operations divisions. Data charts and graphs are introduced by a short description of the various programs, the number of people served, criteria for program participation, and high-level demographic information if available.[5] In subsequent pages, data elements are tracked across an 18-month time span. Attached to each data row is a sparkline, small data lines that present the general shape of variation of the measurement across the span of eighteen months. Seamlessly incorporated into the report, these sparklines serve as "data words" and are intended to be a succinct and memorable representation of the data counts that precede it.

Many of SCHSD's divisions have a dedicated Trendex Edition[6] produced for division-level consumption. Some areas, specifically the WPR (Work Participation Rates) Edition, delve into individual performance measures. The Census Trendex Edition, however, serves as a resource for all users of the tool. Taking data from the decennial census and the American Community Survey, this edition provides an overview of the general characteristics for Sonoma County residents. The population demographics section covers age, ethnicity, citizenship, poverty rates and levels, school enrollment, and disability. Family, household and housing characteristics offer insight into relationship statuses, languages spoken at home, family and household incomes and inflation, and housing types and costs. Employment characteristics are also tracked, looking at the county's labor force and occupation types. All told, this information places SCHSD and the services they provide in the broader context of community.

5. Demographic information includes gender, age, language spoken and ethnicity.

6. Presently, a Trendex edition is being produced for the following divisions: Adult and Aging, Economic Assistance, Human Resources, Work Participation Rates for All Families, and soon-to-come, Employment and Training.

Lastly, the Executive Status Update provides high-level information for the SCHSD executive team. Data elements from significant or especially identified programs are culled from the Main HSD Trendex and presented in this "at-a-glance" iteration that is reviewed at monthly management meetings.

Utilitiy. Trendex is primarily used by management to review historical information and discuss emerging trends. The SCHSD executive team regularly adds Trendex among their meeting agenda items, adding richness to the discussions that drive informed decision-making that adhere to the principles of an upstream approach. Division directors rely on the tool as a central repository for information across the organization, readily available in both hard copy and online. Several directors have uploaded soft copies of the tool to their mobile devices for easy access to data when away from the office, including meetings with county executives and follow up with media inquiries. Gone are the days when directors would guess at information or try to contact others who may or may not know the answers they are seeking.

In several locations, summary graphs taken directly from the reports were made into posters and displayed in high traffic areas. This generous act of sharing department-wide information with the staff that helps collect and record data from which they originate further exemplifies the utility of the reports as an integral tool that drives policy and programs. With Trendex, staff members can confidently speak to the issues and trends facing their client base with the aide of robust data sources at hand. Since its inception, Trendex has become a vital and necessary tool whereupon users depend on information to assist in operations.

Supplements. Consistent among users of Trendex is their unified affirmation that the tool "tells their story," intimating that their history of challenges and successes is documented and depicted within its pages. One way that IID staff have developed to capitalize on this storytelling is the development of periodic "Trendex Supplements," which are designed to increase knowledge and understanding of selected SCHSD program areas through an in-depth drill-down of information. Once a topic area is selected, the team gathers and analyzes data until the analysis tells a compelling story. It is said that the best business decisions come from insights informed by data. Using data in this way allows the organization to build institutional knowledge and creativity on top of a solid foundation of data-driven insights. The findings are then compiled in a presentation format that is initially presented to the executive team and subsequently shared to all SCHSD staff upon posting to their Intranet.

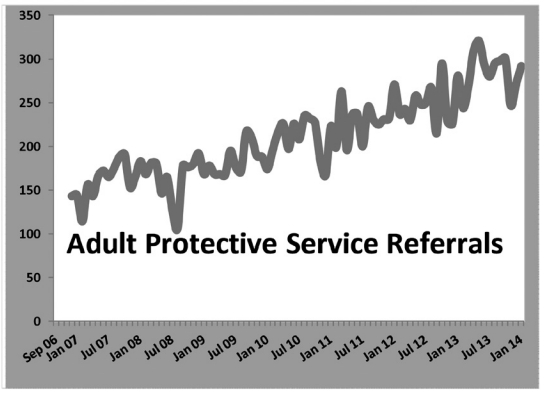
IID staff generally adhere to the following steps in their development of the Trendex Supplement:

1. **Look for data that affect key metrics.** Case scenarios based on real experiences are described. Vivid details related to the topic area are outlined and set the scenario for the study's significance.
2. **Present data so that everyone can grasp the insights.** The resulting PowerPoint presentation focused on telling a clear story with simple slides and visuals. Visualized data are used to find trends, harnessing the power of an image to represent underlying data tables with sometimes ambiguous relationships. By presenting the data visually, those viewing the presentation are able to quickly grasp and contribute to the conversation, which can then lead to a broader insight.
3. **Return to the data with new questions.** Upon reaching a level of understanding about the topic, users can return to the data to ask pointed questions that challenge assumptions, improve processes, tailor services, reach out to those with emerging needs, or highlight potential areas for growth.

Viewing information in this context drives home the concept that **"data gives you the *what*, but humans know the why."**[7] Most data stories

7. Bladt, Jeff & Filbin, Bob. "A Data Scientist's Real Job: Storytelling" (27 March 2013). *Harvard Business Review.*

FIGURE 1
APS Referrals, Sonoma County
(Jan. 2007 – Jan. 2014)



focus on the setting and the plot, but very little on the other key elements of storytelling. In doing so, they miss out on significant opportunities. Take Sonoma County's data on Adult Protective Services (APS) Referrals from January 2007 through January 2014 (*Figure 1*), for example. One can surmise a general trend of increasing referrals (plot) over the course of the past seven years (setting).

However, if the five elements of storytelling are applied to this data story, a more comprehensive and rounded picture emerges (*Figure 2*). When "characters" are assessed, we are informed of the interplay between the public who reports the abuses and the social workers who investigate them. Suddenly, the "conflicts" emerge. The public is becoming increasingly aware of what constitutes elderly abuse, and becomes empowered to report them. On the other hand, the existing social worker pool sees an increase in their investigation caseloads and the inevitable backlog looms to impact services.

The overarching theme that emerges from this story gives SCHSD actionable next steps; namely, to request approval to hire additional resources within APS to address the growing caseloads resulting from increased referrals. In late February 2014, SCHSD received approval from the county executive for three new social worker positions in APS.

FIGURE 2
Application of Storytelling Elements to APS Referral Trends

STORY ELEMENT	KEY QUESTIONS	EXAMPLE
SETTING	When and where is this happening?	APS call centers over the past seven year
PLOT	What is the element of interest?	Significant increases in elderly abuse reporting
CHARACTER	Who is impacting/impacted by the change?	Community/public Social workers
CONFLICT	What business problem does it affect?	Increased public awareness Increased worker caseloads
THEME	What is the opportunity? Where can we go from here?	Justification for request of additional staffing

Data stories continue to get written and rewritten, and Trendex, as a living document, continues to combine data with the five key elements of storytelling, to empower organization leaders in making critical business decisions. Over the course of continuous refinements, and despite some manual processes in collecting data, Sonoma County's Trendex has developed into an information reporting approach that is modeled around the elements that make up an engaging story translated into the visuals and text that make up the presentations. This cross-departmental, in-depth, and hyper-crafted style of data storytelling is difficult, but it is what distinguishes Trendex from standard reports.

Implications for Santa Clara County

Trendex was borne out of a need for a tool that serves as a central repository for information and knowledge to inform the SCHSD about performance and to assist management in the decision-making process. Presently, Santa Clara County Social Services Agency (SCCSSA) has several processes for data collection, analysis, reporting, and distribution. Reports are generated by several teams across many parts of the organization. Information is readily shared across and within departments through electronic repositories that can be subsequently printed and distributed to data users. That said, this patchwork of technical knowledge and data libraries lacks synergy and cohesiveness. Data analysts from the four departments primarily operate as independent teams despite sharing a common skill set and job function. The resulting reports do not share a common composition or design, which makes collaborative work across departments challenging. Some reports do not have a transparent data dictionary, source, or calculation, which makes the replication process difficult. This reliance on individual report builders poses the risk of losing institutional knowledge for the agency. As such, SCCSSA would greatly benefit from investing in the development of a consistent practice for capturing, analyzing, and depicting data, trends, and information.

Recommendations. Santa Clara County Social Services Agency is in a unique position to take lessons learned from Sonoma County to implement is own version of Trendex. While the effort will take an iterative process over several years of refinement, the agency possesses the appropriate resources and tools to begin this transformation. The following cost-neutral recommendations outline the critical next steps for the Agency:

- Consolidate the reporting teams under one Agency-wide structure. The Information Systems Data Services (DS) team currently serves as builders of data universes from which most reports are created. They are viewed as the data experts and primary resource for the genesis of the Agency's business intelligence. It is a natural fit that the DS team plays a leadership role in fostering knowledge sharing among database administrators and report generators. Additionally, as a unified entity, this consolidated unit can better build robust processes that do not rely on individual institutional knowledge, but instead commit to replicable steps and output.
- Leverage the Data Warehouse for the expansion of trends reporting and information access. When SCCSSA invested in the building of the Data Warehouse (DW), it was with expressed intent to bring this tool to all areas of the organization. The warehouse is a stable, read-only database that combines information from separate systems into one easy-to-access location. While a number of departments use the tool to retrieve automatically generated reports, a large portion can still benefit from the value that DW brings. Already accessible online (instead of a shared drive), mobile apps should be developed to enable remote access to the agency's DW, allowing for real-time retrieval of needed information.
- Build performance reports at all levels of the organization. Standardizing on reports built in the warehouse, SCCSSA can replicate and expand upon many of the elements in Sonoma County's Trendex Reports. Call center and

task data can be pulled into the DW and subsequently develop department-wide, office, unit, and individual performance statistics. From this information, management will be better informed about emerging issues and correct for it before they escalate into bigger problems.

Data analysis and storytelling at this level strives to create a clear, more meaningful picture of complex metrics. It is an essential skill to effectively cross the chasm of understanding and ultimately result in action. Charts and tables alone do not necessarily make for an effective vehicle of communicating important findings. Developing a tool and approach to manage organizational information into a coherent narrative turns data into a powerful communication tool.

Acknowledgments

The staff at Sonoma County Human Services Department graciously hosted this case study and provided invaluable insight into Trendex and how it has impacted their business operations and decision-making process. Jerry Dunn, SCHSD Director, was extremely generous with his time, wisdom and enthusiasm for Trendex, and unsparingly offered so much of his staff's ongoing participation to my learning opportunity. A debt of gratitude also goes out to Kiergan Pegg, Diane Kaljian, Gary Fontenot, Kim Seamans, Roy Redlich, Sherry Alderman, George Malachowski, and Fermin Espinoza for sharing their wealth of knowledge. I would like to extend my heartfelt thanks to Tara Smith, Program Planning Evaluation Analyst, for coordinating meetings, preparing materials and resources, and responding to post-exchange inquiries with great aplomb. Finally, I want to recognize Santa Clara County Social Services Agency Director Bruce Wagstaff, Mary Shamouel (Chief Information Officer), Ely Turkenitz (Data Services Director) and Rosemary Semprez (Project Management Director) for their support, encouragement, and the opportunity to participate in BASSC.