

# The Changing Clinical Data Analytics Landscape - Visualization & Beyond

Bhavin Busa, Vita Data Sciences, Waltham, MA, USA  
Siddartha Kondapally, Vita Data Sciences, Waltham, MA, USA  
Kishore Pothuri, Vita Data Sciences, Waltham, MA, USA

## ABSTRACT

“A picture is worth a thousand words” is a commonly used phrase in every industry where there is data involved! In the clinical industry, plotting data and presenting statistical information using graphics is not new. Until most recently, the traditional approach was to “look” at the static output without giving end users an ability to explore and drill-down further. With the rapidly increasing availability of data visualization and analytic tools, the industry landscape is shifting. End users now want to ‘see’ their data more interactively, identify trends, visualize the patient profiles and review results at a high-level while still being able to drill-down to get a complete picture. In this paper, we will talk about how we have used the power of SAS® and TIBCO Spotfire®, to build our “out-of-the-box” dashboards to meet these demands. We will discuss our experiences, processes, case scenarios, and challenges towards implementing such a solution.

## INTRODUCTION

“A picture is worth a thousand words” is a commonly used phrase in every industry where there is data involved! In the clinical industry, the data is represented and analyzed in the form of tables, listings and figures/graphs (TLF or TLG) which are typically generated using SAS®. For the Biometrics teams, plotting data and presenting statistical information using SAS ODS and SAS graphics procedures is not new. However, the output that is generated using SAS ODS and SAS graphics procedure is for the most part static and only gives end users an ability to “look” at the output without giving them an ability to explore and drill-down further.

For any data needs, the clinical study team relies on their Biometrics function (Biostatistics and Statistical Programming) to produce the Clinical Study Report (CSR) TLFs and any data listings/patient profile for their data review and medical monitoring. However, with the rapidly increasing availability of data visualization and analytic tools, the industry landscape is shifting. In addition, the end users now want to ‘see’ their data more interactively, identify trends, visualize the patient profiles and review results at a high-level while still being able to drill-down to get a complete picture. They also wish to have access to their ongoing study data on ‘Day 1’ and don’t want to wait for the study CSR TLFs to be programmed before they can use the data for their exploratory, review or monitoring needs. Although one could argue that the data could be reviewed directly from the EDC system, in our experience, there are many limitations to that approach.

## TOOLS FOR DATA ANALYTICS & VISUALIZATION

There are various options for data analytics and visualization tools in the market today. The top three tools that appear in our list to evaluate were: TIBCO Spotfire®, Tableau®, and Qlik Sense®. After going through extensive market research and referencing community published materials [1, 2, 3, 4], we chose to go with TIBCO Spotfire as our preferred visualization tool because it met most of our needs and requirements. We also continue to use SAS for the purpose of data mapping and transformation needs and to utilize it for the generation of the study CDISC datasets and CSR TLFs as well.

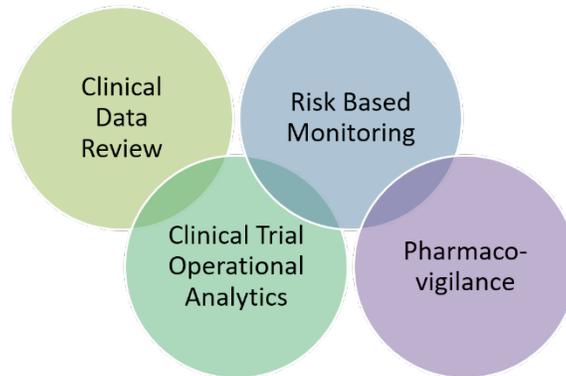
With the combination of SAS and Spotfire, we are able to:

- Continue to leverage our current SAS programming expertise towards data mapping and transformation
- Integrate with the current business process, data standards, and analytics environment
- Utilize advance visualization features to interact, identify, review and visualize study data with ease
- Support publishing of visualization on the web for end users to access via Spotfire Consumer license
- Collectively review and analyze the study data in one central platform
- Export visuals/datasets effortlessly in form of CSV, PDF, or Microsoft PowerPoint
- Navigate between summary level and subject level details
- Generate patient profile, both graphical and in listing format
- Flexibly change analysis elements by using data filters

# PharmaSUG 2018 - Paper DV-18

- Select or 'mark' data and/or subject(s) of interest
- Link the visuals from one page to the other

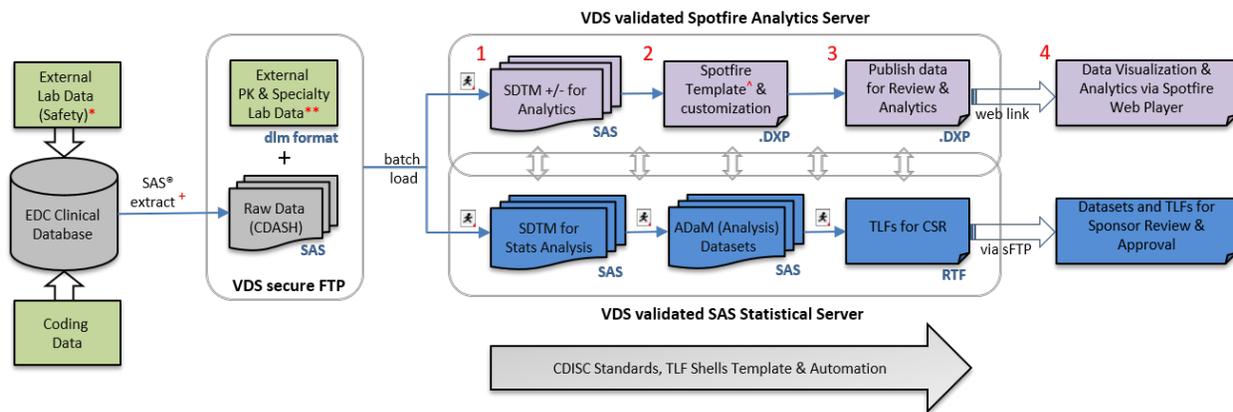
With the power of both the tools, SAS and Spotfire, we at Vita Data Sciences (VDS) are able to build a robust data analytics and visualization solution to provide our clients with an opportunity to drill-down through multiple levels of detail to aid in their clinical data review, operational analytics, risk-based monitoring and safety reporting needs.



**Figure 1:** VDS Data Analytics and Visualization Solution Offerings in Clinical Development via SAS and Spotfire

The solution provided our internal team and our clients with an ability to explore the data for an entire study in a more systematic manner. It is actively helping our clients further assess any potential safety signals, identify dose-response relationships, and identify risk factors in the target population.

## DATA FLOW FOR ANALYTICS/VISUALIZATION & STATISTICAL ANALYSIS



\* Assuming external lab safety data and coding data will be integrated into the EDC system.  
 \*\* External specialty data could be integrated into the EDC system. In this case, raw data extract will be inclusive of this data and SDTM will be programmed based on extract from the EDC directly.  
 † Scheduled transfer (via automation services).  
 ‡ VDS developed standard dashboard template via existing Spotfire library.

SAS-based standardized + study-specific scripts/macros

Connection across both SAS Statistical & Spotfire Analytics Servers

**Figure 2:** A schematic representing high level flow of data for Data Analytics & Visualization (shaded in purple) and Statistical Analysis Deliverables (shaded in blue) deployed at VDS

The Figure 2 above provides a high-level flow for data analytics & visualization (shaded in purple) along with the statistical analysis deliverables (shaded in blue). For the most part, the steps are self-explanatory, however, we wanted to provide further details into the analytics components annotated with numbers in red fonts. Once the raw data and external data (if applicable) is available on our analytics server via an automated batch load, the steps for data analytics are initiated as follows:

Step 1: SDTM +/- will be generated via SAS standard and study-specific scripts. VDS has existing suite of in-house

# PharmaSUG 2018 - Paper DV-18

validated macros to generate domains required for analytics needs. The SDTM datasets generated in this step will be programmed per the data analytics needs, "+" indicating addition of certain derived fields which will facilitate data review (e.g. change from baseline, flagging of lab values per pre-defined criteria's, sub-group variables), while "-" indicating not adding domains/variables that are not required towards data review.

Step 2: In this step, the SDTM +/- data will be loaded into study specific instance of Spotfire dashboards and templates. The study specific instance will include VDS "out-of-the-box" standard global templates along with the customized reports per the user requirements. A validated DXP file(s) will be generated which will be utilized throughout the study and can be re-used across future studies.

Step 3: In this step upon validation of dashboards & reports, data will be published on the Spotfire servers for review by the Sponsor team. A web link will be shared with the Sponsor team for access to the data. An auto-email communication can be set-up to notify study team of new data availability or to notify out-of-range data values.

Step 4: Sponsor team will utilize Spotfire Web Player/Consumer instance to visualize and review data during the study.

## DATA VISUALIZATION VIA SPOTFIRE

The VDS "out-of-the-box" data analytics solution for Clinical Data Review via Spotfire provides our clients the ability to use standard dashboards and global templates from Day 1. The dashboard provides them with an ability to start their clinical data review and visualization process very early during the trial. This "out-of-the-box" solution includes templates that minimally cover data review/visualization and workflows of Patient Demographics, Enrollment over time, Enrollment by Site and Country, Adverse Events, Labs (LFT plot, Box plot, Over-time plot), Vitals (Scatter plot, Box plot, and Over-time plot), ECG, Concomitant Meds, and Patient Profile (Graphical and Listing formats).

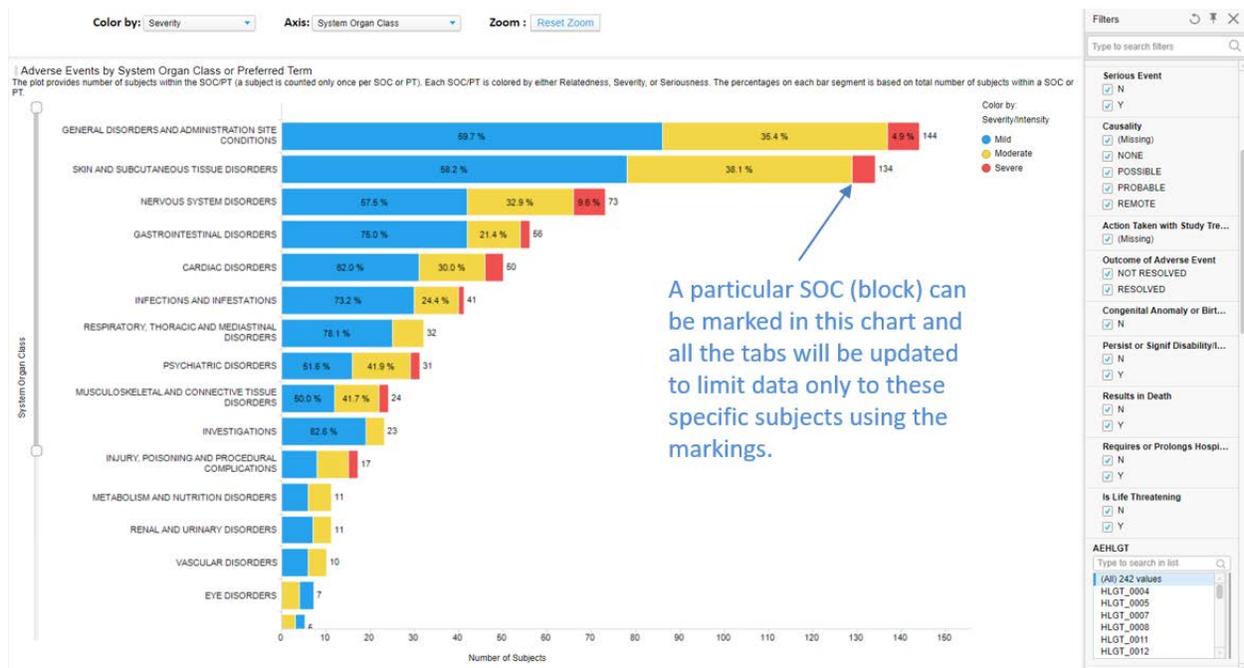


Figure 3: Screen shot of the Adverse Events by SOC page from our VDS Data Analytic instance - annotated to show how one can mark/select any part of the graph by clicking on it

The Figure 3 above provides a view of Adverse Events (AE) by System Organ Class (SOC) and is colored by AE Severity. The same page could be used to display either SOC or PT and can be colored by either AE Severity, Seriousness or Relatedness. The percentages on each bar segment is based on total number of subjects within a SOC or PT. In Spotfire, a marking (i.e. selection) can be made in the chart which will update all the pages/tabs which are linked to this visual to limit data only to these specific subjects.

The columns in the filter section can be used for sub-setting data and review the data at granular level. Pre-defined data filters present to the right are to only display value of interest. When one uses a filter, certain data values in the data table are filtered out, and only the filtered values, that is the remaining values, are included in what is being visualized. All visualizations that are based on this data table will immediately update to reflect only the filtered

# PharmaSUG 2018 - Paper DV-18

values. However, you can anytime change what is filtered, or return to the full set of data [5, 6].

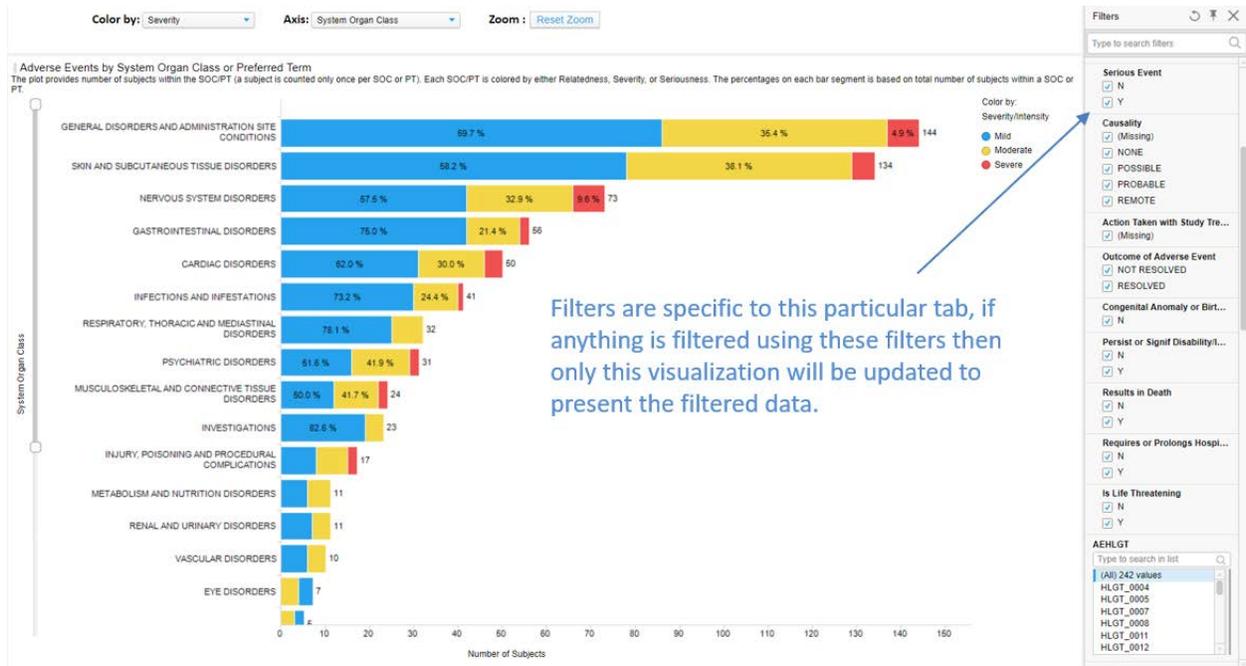


Figure 4: Screen shot of the Adverse Events by SOC page from our VDS Data Analytic instance - annotated to show how one can select pre-defined data filters to the right to only display value of interest

The Figure 5 below shows how the same graph is updated to only display Serious AEs based on the data filter selected (Serious Event = "Y"). The user is quickly able to tell that there are 3 Serious AEs reported in this study and all SAEs fall under the same SOC of "Nervous System Disorder". In addition, the user will be able to know 2 out of 3 (66.7%) is reported as Severe, while 1 out of 3 (33.3%) is reported as Mild. Further, the user can change the Y-axis to look at the Preferred Term associated with these SAEs and can also understand the causality by changing the Color by drop-down to Relatedness. One can also use multiple filtering condition to restrict data per their needs and the visual on the current page will be updated accordingly.



Figure 5: Screen shot of the Adverse Events by SOC page from our VDS Data Analytic instance - restricted to display Serious AE based on the selection of a data filter Serious Event = 'Y'

## PharmaSUG 2018 - Paper DV-18

The source of the data that is being used for the visual can be viewed in the Details-on-Demand area. Figure 6 shows the Details-on-Demand panel presenting the underlying source data of the visual in Figure 5. When a particular data is marked in the visual then all the rows of the data of that part will be presented in the Details-on-Demand panel. This functionality can also be used to drill down into specific details and the data rows can be exported to excel sheet. In this example, a data listing of 3 SAEs are presented which can be exported into CSV for internal review and communication.

Study Site Id...	Unique Subject Id...	Description of Plans...	Reported Ter...	Dictionary-Derived T...	Body System or Organ C...	Serious Event	Action Taken ...	Causality	Outcome of Ad...	Start Date/Time of A...	End Date/Time o...	Study Day of End...
709	01-709-1424	Xanomeline High Dose	VERBATIM_1...	SYNCOPE	NERVOUS SYSTEM DIS...	Y	(Missing)	POSSIBLE	RESOLVED	2013-03-07	2013-03-07	5.00
718	01-718-1170	Xanomeline Low Dose	VERBATIM_1...	SYNCOPE	NERVOUS SYSTEM DIS...	Y	(Missing)	PROBABLE	RESOLVED	2013-10-12	2013-10-13	29.00
718	01-718-1371	Xanomeline High Dose	VERBATIM_0...	PARTIAL SEIZURES	NERVOUS SYSTEM DIS...	Y	(Missing)	NONE	RESOLVED	2013-06-02	2013-06-05	41.00

**Figure 6:** Screen shot of the Details-on-Demand window displaying underlying source data based on the marking on the Adverse Events by SOC page - restricted to display Serious AE based on the selection of a data filter

## SPOTFIRE CONSUMER/WEB PLAYER

The Spotfire analysis file (DXP file) can be published onto a Web player, where the user can explore data using interactive visualizations. Spotfire Consumer (previously called as a Web Player) is a powerful tool for sharing and distributing information within and outside an organization. It is a web browser-based application, which requires no installation, and can be used inside a portal or a web site [5, 6]. Web player lets you to filter data, drill down to details, identify spots or trends and export data.

## KEY RECOMMENDATIONS AND LESSONS LEARNED

Based on our experience and several case studies, we have summarized some of the key recommendations and lessons learned:

- Challenge the status quo of doing things the traditional way and educate stakeholders about the technology solutions as proposed in this paper that can help study teams make their data review and monitoring activities faster.
- Implementation of any new tool in the organization requires upfront investment to set-up the infrastructure, and processes. The organization (vendor or sponsor) will need to build a dedicated team and resources to support their data analytics needs.
- Develop and maintain standard requirements, global templates and workflows for data visualization (“out-of-the-box” solution). This will be key to maintain consistency and make the process more efficient/cost-effective. The team can re-use the templates from study to study and client to client. Of course, anticipate customization for each study/client and ensure additional effort/budget is built-in to the project plan to cover the study-specific requirements.
- In our proposed data flow, we are recommending the use of SDTM +/- for data analytics and visualization; this requires additional programming efforts. However, given we typically also work on generating SDTM datasets for the study, the programs can be used for both Data Analytics and Statistical Analysis purposes. Utilizing SDTM will result in reusability, faster ramp-up time, and will maintain consistency.
- Although the process could be automated, we do recommend adding a layer of QC upon each refresh of the data for an on-going study. In addition, if the source data is mapped or transformed, the programming validation and QC process should be followed.
- Ensure all end-users are trained on each study-specific data visualization instance. Add a disclaimer that the data represented in tool such as Spotfire is for the data review, exploratory and monitoring purpose only. Data Analytics tools should not be used for any regulatory and/or publication purposes.

## CONCLUSION

In this paper, we talked about how the clinical data analytics landscape is changing. The clinical study team members want to ‘see’ their data more interactively, identify trends, visualize the patient profiles and review results at a high-level while still being able to drill-down to get a complete picture. They also wish to have access to their ongoing study data on ‘Day 1’ and don’t want to wait for the study CSR TLFs to be programmed before they can use the data for their exploratory, review or monitoring needs. We think with the use of SAS and Spotfire, we are able to provide a solution that meet these needs.

We have covered some of the examples in the above sections. However, there are many more features and functionalities built-in to our “out-of-the-box” solution. This solution provides our clients and the study team with an ability to drill-down to get a complete picture of an ‘outlier’ or ‘data of interest’ by using our marking and filtering functionality. In addition, the user can review and analyze the data collaboratively by publishing the results to the

## PharmaSUG 2018 - Paper DV-18

study team members. Moreover, we can utilize the collaboration tool within Spotfire to comment on the analysis state or any specific data points. Lastly, the user is able to export any visuals or underlying source datasets effortlessly in form of CSV, PDF, or Microsoft PowerPoint.

With the power of SAS and Spotfire, we were able to build a solution where user is able to seamlessly navigate between summary level and subject level details by reviewing the built-in patient profile pages, both graphical and in listing format. At VDS, we have built multiple dashboards that support various clients towards their data analytics and visualization needs.

### REFERENCES

- [1] Spotfire and Tableau Positioning paper, by Martin Butler.  
<http://marketo.tibco.com/rs/221-BCQ-142/images/martin-butler-spotfire-and-tableau-positioning.pdf>
- [2] SAS Advanced Analytics vs. TIBCO Spotfire.  
<https://www.trustradius.com/compare-products/sas-advanced-analytics-vs-tibco-spotfire>
- [3] Spotfire and Qlik Sense Positioning White Paper  
<https://spotfire.tibco.com/resources/whitepaper/spotfire-and-qlik-sense-positioning-white-paper>
- [4] How to Choose Between Tableau and Spotfire Analytics for Your Company, by Butler Analytics  
<https://spotfire.tibco.com/resources/whitepaper/how-choose-between-tableau-and-spotfire-analytics-your-company-butler>
- [5] TIBCO Spotfire® Business Author and Consumer User's Guide.  
<https://docs.tibco.com/pub/sfire-consumer/7.9.0/doc/html/en-US/GUID-6CF74B7C-9425-4BDE-B8C4-F9F45B1F1710.html>
- [6] TIBCO Spotfire® Business Author and Consumer User's Guide.  
[https://docs.tibco.com/pub/spotfire\\_web\\_player/6.0.0-november-2013/WebPlayerUserGuide/GUID-2813F3F3-A2D5-4868-8B99-ABC358E5A41C.html](https://docs.tibco.com/pub/spotfire_web_player/6.0.0-november-2013/WebPlayerUserGuide/GUID-2813F3F3-A2D5-4868-8B99-ABC358E5A41C.html)

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Bhavin Busa  
Company: Vita Data Sciences, a division of Softworld, Inc.  
Address: 281 Winter St, Waltham, MA 02451  
Work Phone: 781-373-8455  
Email: [bbusa@vitadatasciences.com](mailto:bbusa@vitadatasciences.com)  
Web: [www.vitadatasciences.com](http://www.vitadatasciences.com)

Brand and product names are trademarks of their respective companies.