

# STD Analysis and Visualization Tool (SAViT)

**Rodney Presley**

CDC/NCHHSTP/DSTD Data Management Team Lead

**Massimo Mirabito**

NCHHSTP Informatics & Northrop Grumman

CDC Health Information Consortium (CHIIC)

November 8, 2016

## Contributors and Acknowledgements

### CDC

- *Alesia Harvey*
- *Doug Correll*
- *Elizabeth Torrone*
- *Craig Cousins*
- *Hillard Weinstock*
- *Robert Nelson*
- *Thomas Sukalac*

### Northrop Grumman

- *Kumar Batra*
- *Rishi Tarar*
- *Silver Wang*

# Presentation Overview

- **Current tool – STD Net**
  - Capabilities
  - Limitations
  - Desirable capabilities of replacement tool
- **STD Analysis and Visualization Tools (SAViT) project**
  - Findings and recommendation
- **Next steps**

# STD Net

## ■ Data sources

- Historical hard copy forms
  - Aggregated by some combination of sex, age, or race
  - Aggregation varies across diseases and years
  - Frequency of reporting varies by years, disease, and geographic unit
  - Some data available from 1963, most from 1981
- Electronic data -- National Electronic Telecommunications System for Surveillance (NETSS)
  - Individual case reports
  - Reported weekly
  - Different jurisdictions adopted NETSS in different years
  - Data available from 1992

# STD Net

- **Capabilities**
  - Calculates case count and rates by multiple combinations of year, area, sex, age, and race for reportable STDs
- **Available outputs**
  - Table
  - .csv file
  - Bar graph
  - Line graph
  - Choropleth map

# STD Net

## ▪ **Limitations**

- Can only select pre programmed queries
- Slow processing speed for queries accessing non aggregated, case level data
- Cumbersome to maintain programs
- Not currently programmed to access SQL data tables

## STD Net

- **Desirable capabilities of replacement tool**
  - Allow user to construct any query
  - Speed up queries of case level data
  - Provide multiple output formats in multiple file formats
  - Provide transparent access to SQL data repository
  - Reduce maintenance and modification efforts

# SAViT

## **SAViT Overview**

- **Analysis conducted between March 2015 and July 2015**
- **High-level assessment of 12 technologies**
- **BI platform vendors assisted and validated the capability matrix**
- **Analysis was not validated or peer-reviewed**

## SAViT Evaluation Criteria

- Architecture - OS requirements, API, Data Access
- Security - Type of authentication (SSL, AD, Kerberos, Oauth,
- Development - IDE, Report Designers, Programming Model, 508
- Visualization - Tabular, Maps, cross-tabs,, drill down, export, Ad-Hoc
- Technical Support - Telephone, eMail, SLAs, forums
- Other - In use at CDC, Open Source, Licensing, Cost

# Detail Evaluation Criteria

<b>► Architecture</b>		<b>► Development</b>		<b>► Technical Support</b>	
Version		Integrated Development Environment	✓	Telephone Support	
Operating Systems		Dashboard designer	✓	eMail Support	
Technology Stack Support		Full programmatic control/scripting	✓	Service Level Agreement	
Requires Plugin (Flash , Silverlight etc)		Grouping, aggregations , sub-totals , calculations	✓	Documentation	
HTML Based Visualization	✓	Multi-section and sub-reports	✓	Web Forums and other resources	
Javascript API	✓	Conditional sections and logic	✓		
REST API	✓	On demand reporting	✓	<b>► Other</b>	
Web Access	✓	Scheduling reports		Currently in use at CDC	
Server Edition		Mobile Ready	✓	Open Source	
<b>Data Integration</b>		Big Data Capable		Source Code Available	
<b>Data Access Support</b>		Section 508 Capable		Multi-User Support	
Multi-Vendor DB Support	✓	<b>► Visualization Capabilities</b>		License	
Multi-Dimensional databases (Data Cube)	✓	Tabular data	✓	Pricing	
ODBC		Map	✓	<b>► Security</b>	
JDBC		Charts	✓	Secure socket Layer Support	✓
Multi-datasource capability	✓	Cross-Tab reports	✓	Active Directory	✓
<b>Data Source Support</b>		Drill Down	✓	SAML - 3rd party integration	
Microsoft Excel File	✓	Dash-Boards	✓	Kerberos	
Flat Files	✓	Export Excel	✓	Oauth 2.0	
XML Files	✓	Export PDF			
Scripted Datasource	✓	Export Word			
Statistical files (SAS, STATA, R)	✓	Export Powerpoint			
		Ad-Hoc Reporting	✓		
		Ad-Hoc Analysis			

# Software Evaluated

## BI Platforms

- Actuate BIRT
- BIRT (Open Source)\*
- Cognos
- InstantAtlas\*
- Pentaho Premium
- Pentaho (Open Source)
- QlikView
- SAP Business Objects\*
- SAS Visual analytics\*
- Tableau\*

## Visualization Libraries

- D3.JS\*
- FusionCharts\*

## Not Evaluated

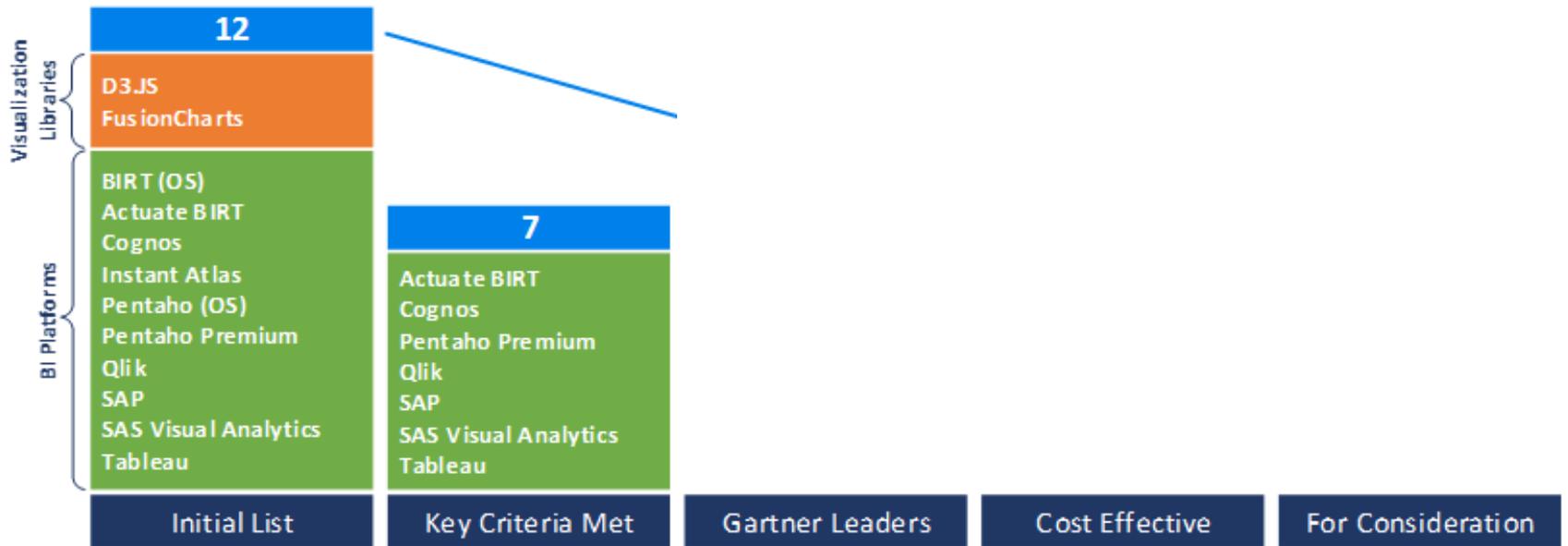
- Microsoft
- Palantir\*

*\*Currently at CDC*

# Software Analysis

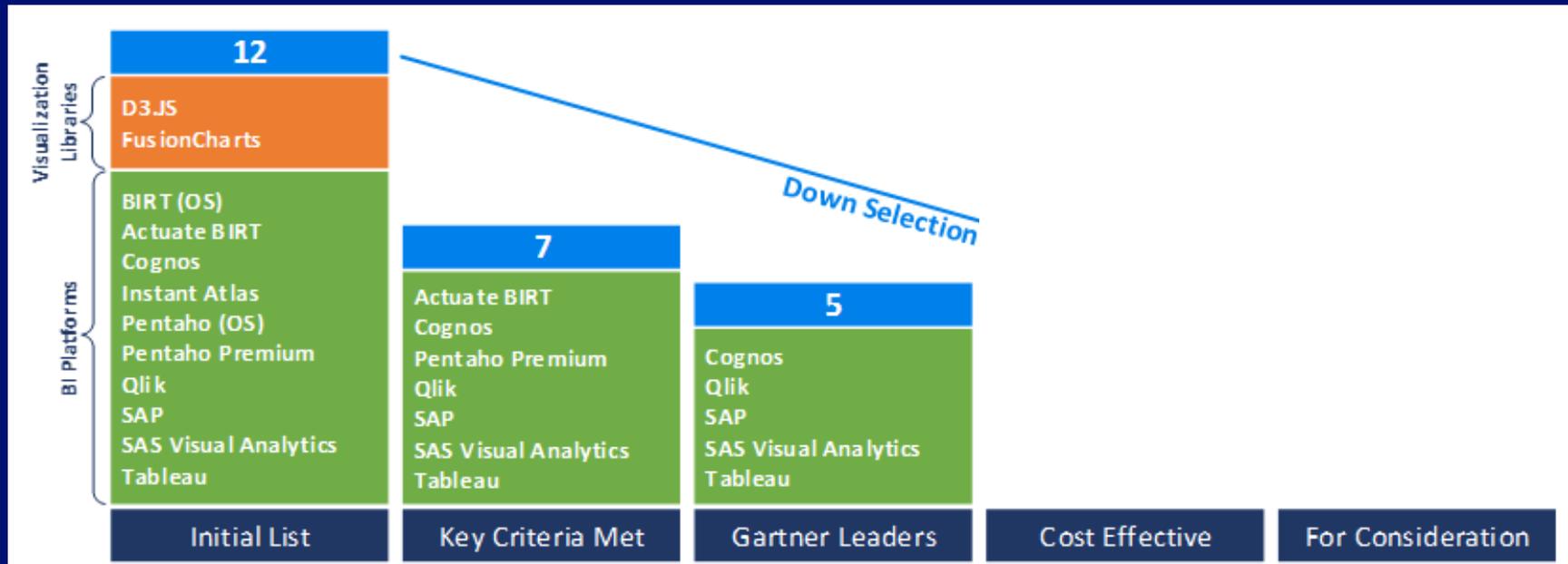


# Software Analysis (continued)

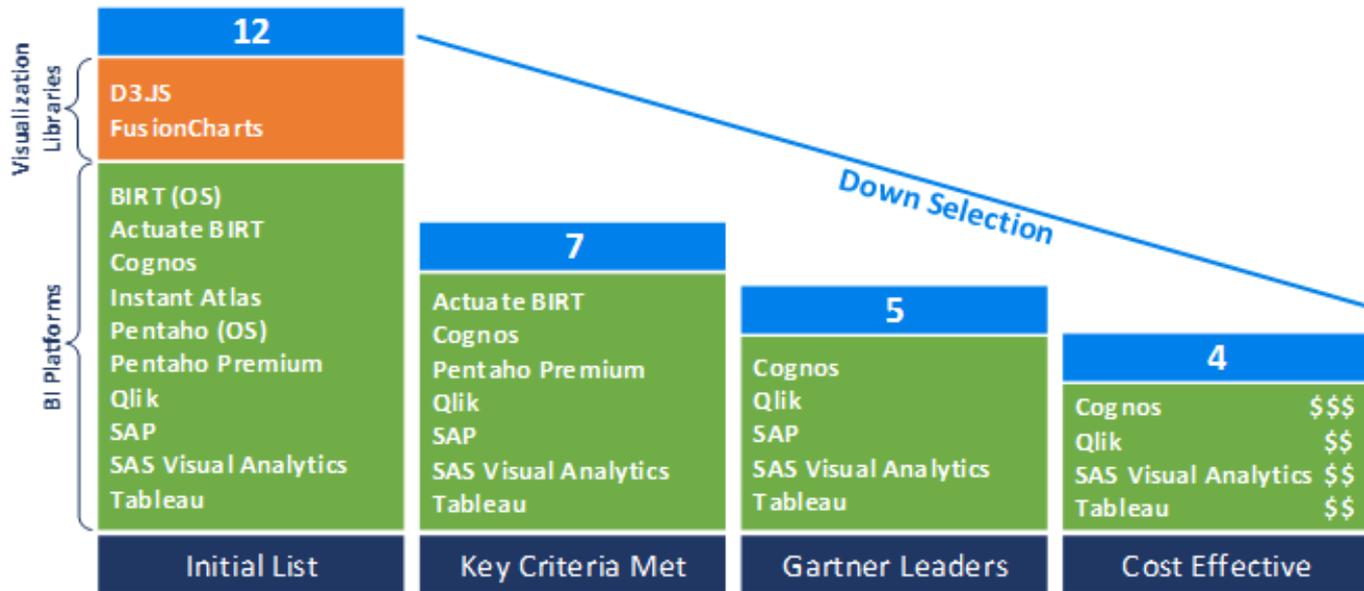


<b>► Architecture</b>		<b>► Development</b>	
HTML Based Visualization	✓	Integrated Development Environment	✓
Javascript API	✓	Dashboard designer	✓
REST API	✓	Full programmatic control/scripting	✓
Web Access	✓	Grouping, aggregations, sub-totals, calculations	✓
<b>Data Integration</b>		Multi-section and sub-reports	✓
<b>Data Access Support</b>		Conditional sections and logic	✓
Multi-Vendor DB Support	✓	On demand reporting	✓
Multi-Dimensional databases (Data Cube)	✓	Mobile Ready	✓
Multi-datasource capability	✓		
<b>Data Source Support</b>		<b>► Visualization Capabilities</b>	
Microsoft Excel File	✓	Tabular data	✓
Flat Files	✓	Map	✓
XML Files	✓	Charts	✓
Scripted Datasource	✓	Cross-Tab reports	✓
Statistical files (SAS, STATA, R)	✓	Drill Down	✓
<b>► Security</b>		Dash-Boards	✓
Secure socket Layer Support	✓	Export Excel	✓
Active Directory	✓	Ad-Hoc Reporting	✓

# Software Analysis (continued)



# Software Analysis (continued)



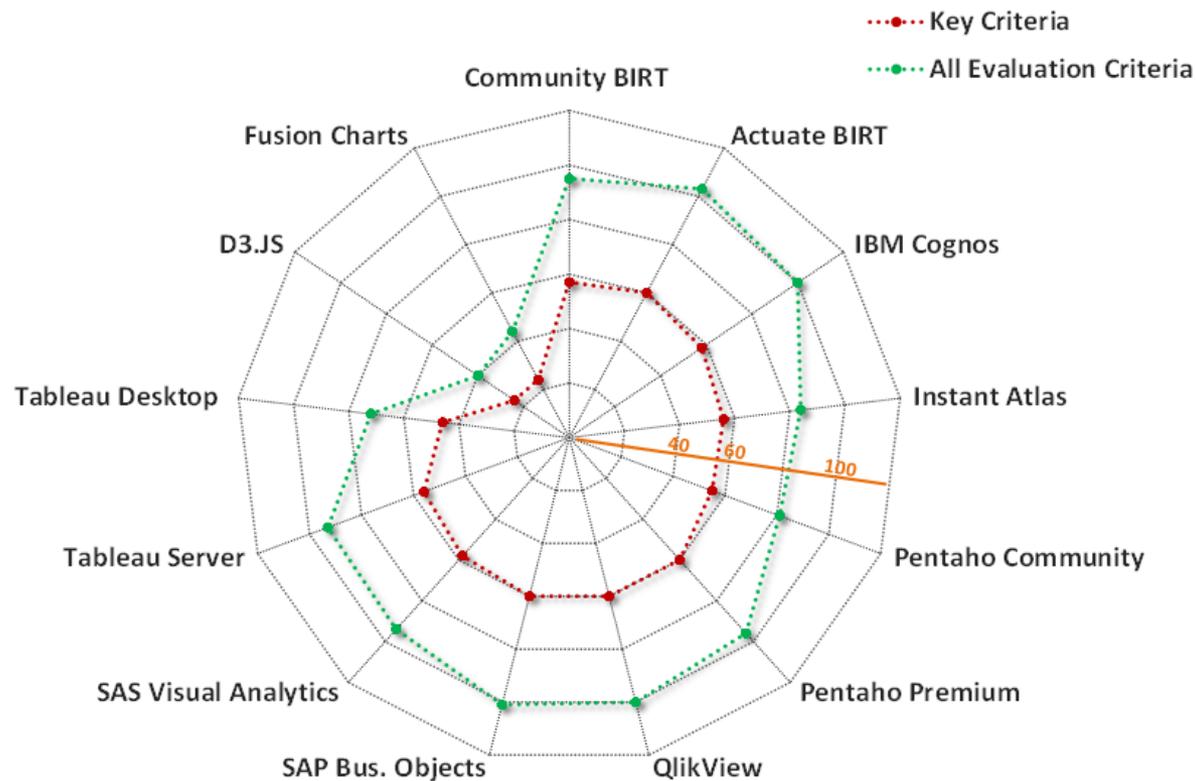
	Licensing Type	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	TCO
<b>Actuate</b>	Subscription	\$ 156,000	\$ 196,000	\$ 196,000	\$ 196,000	\$ 196,000	\$ 196,000	\$ <b>817,267</b>
<b>Cognos</b>	User Licencing	\$ 111,404	\$ 58,564	\$ 58,564	\$ 58,564	\$ 58,564	\$ 58,564	\$ <b>303,098</b>
<b>Qlik</b>	User Licensing	\$ 79,000	\$ 55,800	\$ 55,800	\$ 55,800	\$ 55,800	\$ 55,800	\$ <b>264,114</b>
<b>SAP</b>	Server Licensing	\$ 534,692	\$ 136,419	\$ 147,333	\$ 159,119	\$ 171,849	\$ 185,597	\$ <b>1,029,669</b>
<b>*SAS Visual Analytics</b>								
<b>Tableau</b>	Named users	\$ 55,630	\$ 53,930	\$ 53,930	\$ 53,930	\$ 53,930	\$ 53,930	\$ <b>236,425</b>
	Server Licensing	\$ 273,937	\$ 108,494	\$ 108,494	\$ 108,494	\$ 108,494	\$ 108,494	\$ <b>622,922</b>

\* SAS Visual Analytics will be deployed and maintained by ITSO and DSTDP will not incur any infrastructure cost. Annual licensing costs are still valid and will be equivalent to the negotiated price between the CDC and SAS and will be captured through the Working Capital Fund (WCF). SAS projections are based on the current rate of \$1,172 per user and will be renegotiated in 2018.



# Software Analysis

Comparison by All Criteria vs. Key Criteria



# Software Analysis (continued)

ACTUATE The BIRT Company™			COGNOS®			InstantAtlas™ visualize   communicate   ENGAGE			pentaho			QlikView		
HTML	<>	DB	HTML	<>	DB	HTML	<>	DB	HTML	<>	DB	HTML	<>	DB
Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar
Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile
SAP			sas			+ a b l e a u			Data Driven Documents			FusionCharts		
HTML	<>	DB	HTML	<>	DB	HTML	<>	DB	HTML	<>	DB	HTML	<>	DB
Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar	Doc	Sec	Bar
Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile	Dash	Ad-Hoc	Mobile

Full HTML Stack Support	Full IDE Built-in Support	Multi-Database support	Full support
Support for Data Export	Security Support (RBA)	Data Visualization	Partial support
Dashboard Support	Ad-Hoc Analysis	Mobile Ready	Unsupported

## Key Findings

- The market is saturated with Data Visualization software
- Some products are more mature and capable than others
- Need to clearly define goals and expectations
- There is significant upfront commitment
- BI Platforms require full long-term commitment
- Pricing models may vary
- No single offering to address all needs
- No additional cost to use SAS Visual Analytics until 2018

## Next steps

- **Gather cost information for two contenders**
  - Purchase costs
  - Maintenance cost
  - Personnel costs
  - Infrastructure costs (servers, web portals etc.)
- **Request trial evaluation of tool**
  - Recruit NON data managers / analysts to evaluate

## Next steps

- **Compare notes with other CDC users of data visualization tools**
  - Customer support / service
  - Total costs
  - User acceptance and adoption



# Questions

**For more information please contact Centers for Disease Control and Prevention**

1600 Clifton Road NE, Atlanta, GA 30333

Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

Visit: [www.cdc.gov](http://www.cdc.gov) | Contact CDC at: 1-800-CDC-INFO or [www.cdc.gov/info](http://www.cdc.gov/info)

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention

Division Name in this space

