Making It Happen: A novel way to save taxpayer dollars by implementing an in-house SAS® Data Analytics and Research Center (DAARC)

Vivek Sethunatesan, Technical Advisor – Health IT, Northrop Grumman, Baltimore, MD

ABSTRACT

As part of promoting a data-driven culture and data analytics modernization at its Federal sector clientele, Northrop Grumman developed a framework for designing and implementing an in-house DAARC using an SAS® set of tools. This DAARC provides a complete set of SAS® Enterprise Business Intelligence (BI) and SAS® Data Management tools. The platform can be used for data research, evaluations, and analysis and reviews by federal agencies such as the Social Security Administration (SSA), the Center for Medicare and Medicaid Services (CMS), and others. DAARC architecture is based on an SAS data analytics platform with newer capabilities of data mining, forecasting, visual analytics, and data integration using SAS® BI. These capabilities enable developers, researchers, and analysts to explore big data sets with varied data sources, create predictive models, and perform advanced analytics including forecasting, anomaly detection, use of dashboards, and creating online reports. The DAARC framework that Northrop Grumman developed enables agencies to implement a self-sufficient "analytics as a service" approach to meet their business goals by making informed and proactive data-driven decisions.

This paper provides a detailed approach to how the DAARC framework was established in strong partnership with Federal customers of Northrop Grumman. This paper also discusses the best practices that were adopted for implementing specific business use cases in order to save taxpayer dollars through many research-related analytical and statistical initiatives that continue to use this platform.

INTRODUCTION

Evidence-based research approach is a fundamental part of the Presidents' legislative agenda (second term), No President have ever been so intent on using evidence to shape decisions about the funding of social programs as President Obama. Basing program decisions on rigorous evidence is expected to be useful for cutting spending as well as funding new programs. Data driven decision is changing how Federal agencies are operating today. The field of research-based policy changes to social programs based on data analytics is progressing at a rate beyond anything we have ever experienced. Tapping into this new wealth of information, and making decisions based on it, is helping to transform the way our agencies operate. Evidence-based research approach is aimed at taking the guesswork out of decision making in meeting agency level performance goals. It has the potential to improve our approach to everything from transparency at CMS, where leaders strive to promote data openness, extends to the public. Data on Medicare-funded nursing homes on its Nursing Home Compare website is not only open to staff, delivery partners and stakeholders; but Medicare recipients can compare nursing home facilities to SSA use of mission analytics to Improve access to services by increasing the number of citizens who complete their business with us online.

Northrop Grumman has been working for years with many of its Federal clients to take their deluge of data and help them to make sense of it. During these engagements, one of our key observations is the need for an in-house DAARC. This paper discusses in detail how DAARC provides a platform using SAS Data Management (DM) and Enterprise Business Intelligence (EBI) tools to promote data analytics and data science as a service to assist agencies in making business decisions by enabling users to analyze huge volumes of transaction data and other untapped data sources.

DAARC offerings can include data mining, predictive analytics, applied analytics and statistics and other data analytics and data science services. These techniques are used to build analysis models and simulations to uncover hidden patterns and previously unknown correlations in order to comprehend fundamental business realities and be able to predict future states of the agency and its processes.

This paper assumes the reader has a basic understanding of Medicare and Social security programs, data analytics and data science terminology; base SAS® programming; SAS® Enterprise DM and BI tools;

the Relational Database Management System (RDBMS); the Hadoop data file system platform; and protocols to access data from Hadoop and RDBMS.

WHAT IS DAARC? WHY DO YOU NEED A DAARC?

Agency leaders play a critical role in communicating a clear vision, setting expectations and calling for accountability for results. Agencies don't have to start an analytics program on a grand scale. They can begin benefiting from using an in-house DAARC and grow their programs as they gain experience in performance management.

DAARC is a centralized group of highly skilled resources that offer a wide spectrum of data analytics services accessing authoritative data sources loaded to Hadoop Data Lake platform, using SAS EBI and DM tools and technologies.

DAARC can help agencies derive value-added insights from its data assets and promote innovation through advanced analytics. These analytics will drive strategic objectives, improve organizational performance, and deliver operational excellence. DAARC plays a key role in nurturing and promoting a culture of evidence-based decision-making and policies across the agency by increasing analytics competency. At the same time, it can collaborate with business owners and technology experts to perform insightful analyses of agency data and deliver actionable analytics.

WHO CAN BENEFIT FROM A DAARC?

- Established Analytics teams, not yet able to identify business problems, data sets, or resources to initiate project

- Leaders interested in applying Data Analytics and Data Science to address specific Use Cases

- Agency components with business problems that require in-depth data analyses

- Practitioners that are currently performing analytics in agency components but need guidance

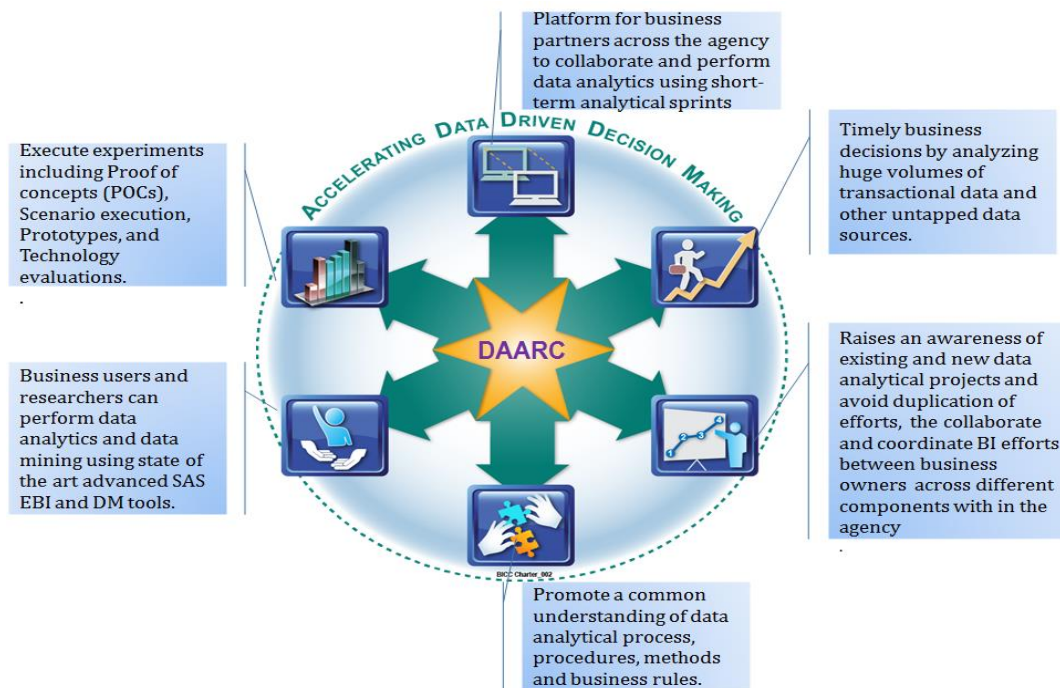- White House, Congress, States, Partner agencies (e.g. OMB, CMS, HHS, VA) and Research Institutes

Figure 1 : DAARC benefits

DAARC REFERENCE ARCHITECTURE

DAARC reference architecture is designed to maximize the full potential of agency's vast data assets and analytical tools and techniques that can support the internal and external analytics for current and future business model and to help accelerate data driven decision making. This reference architecture is carefully designed to improve agency's ability to report facts, analyze trends, and to predict outcomes through rapid analytical sprints. DAARC uses SAS EBI; DM tools to provide a platform for integrate agency data sources to service business intelligence (BI), data analytics and research analytics across the agency.
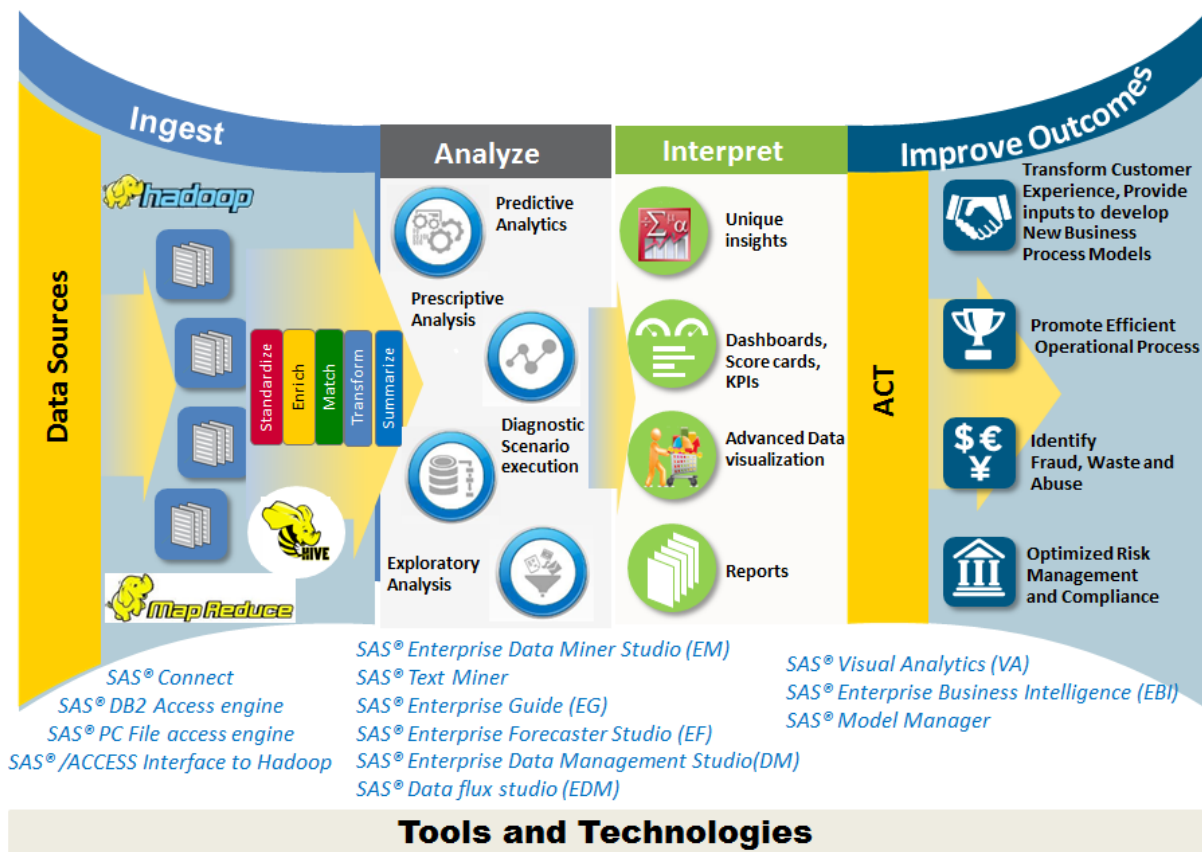

Figure 2: DAARC Reference Architecture (Conceptual)

DAARC reference architecture addresses a variety of uses cases ranging from business intelligence which consumes stored information to Business analytics which produces new information. DAARC analytical projects and research experiments include regression, segmentation, correlation, and clustering analysis. One of the major areas of focus for DAARC is to develop analytical applications that can determine characteristics and patterns of fraud based on data from past allegations and known cases of fraud. DAARC predictive models can support key agency program integrity functions and to help facilitate efficient case/claim processing. These predictive models are estimated using case or claim level data (atomic-data) from agency's programmatic systems which can be many millions of records with known outcomes of the event being predicted.

DAARC DATA-DRIVEN SERVICES

DAARC services will vary on a project-by-project basis, and on the needs and analytics maturity of the requesting user segment.
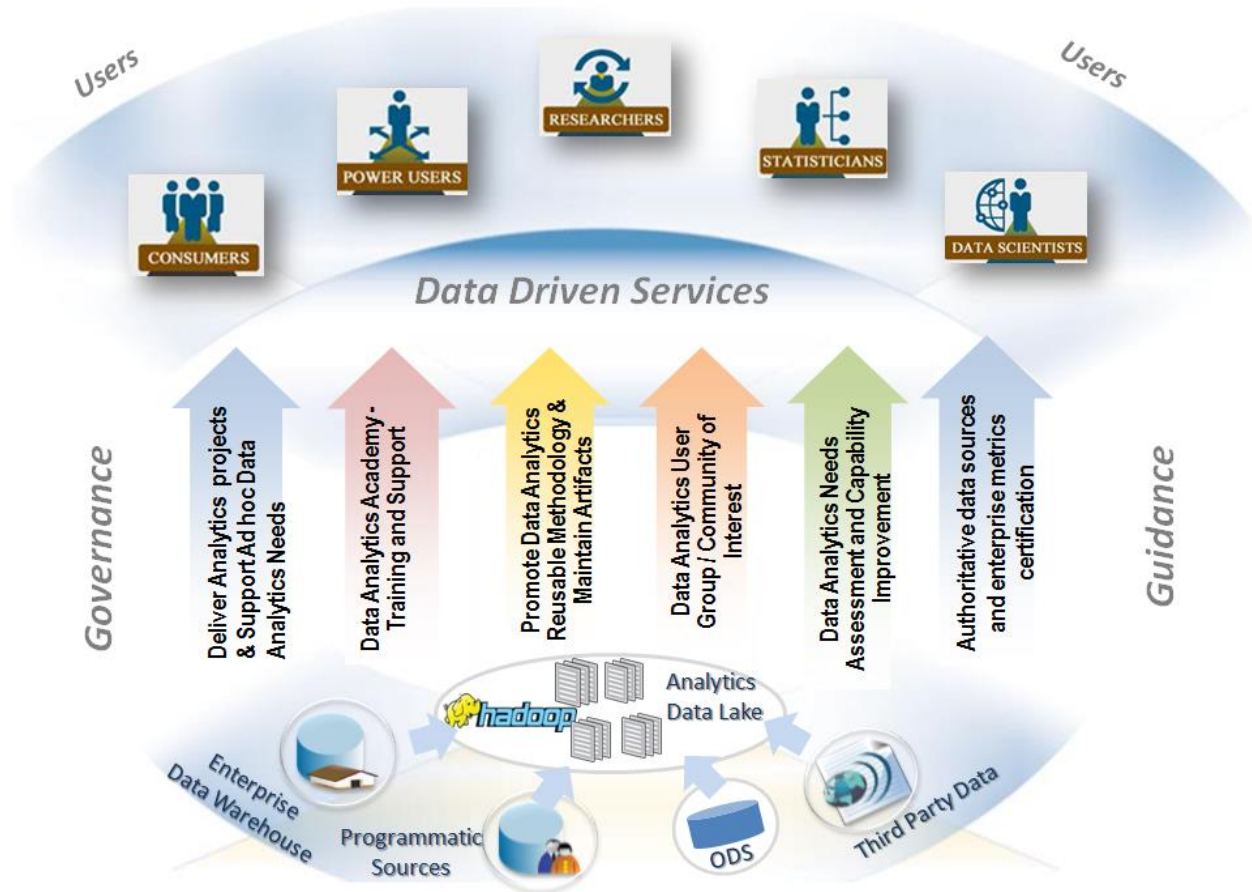

Figure 3 : DAARC data driven services

DATA ANALYTICS – PROJECT LEVEL OR AD HOC

DAARC will undertake advanced analytics initiatives that require in-depth data analyses, use of sophisticated analytics tools and techniques. It will deliver analytics projects that support improvements to operating efficiency and effectiveness, strategic decision-making, and business processes improvements it provides ad hoc data analytics support for policy research, projects, initiatives, and deep dives. Enables Peer Review of analytics techniques and methodology. DAARC involves one of following types of analytics as shown in the figure below.
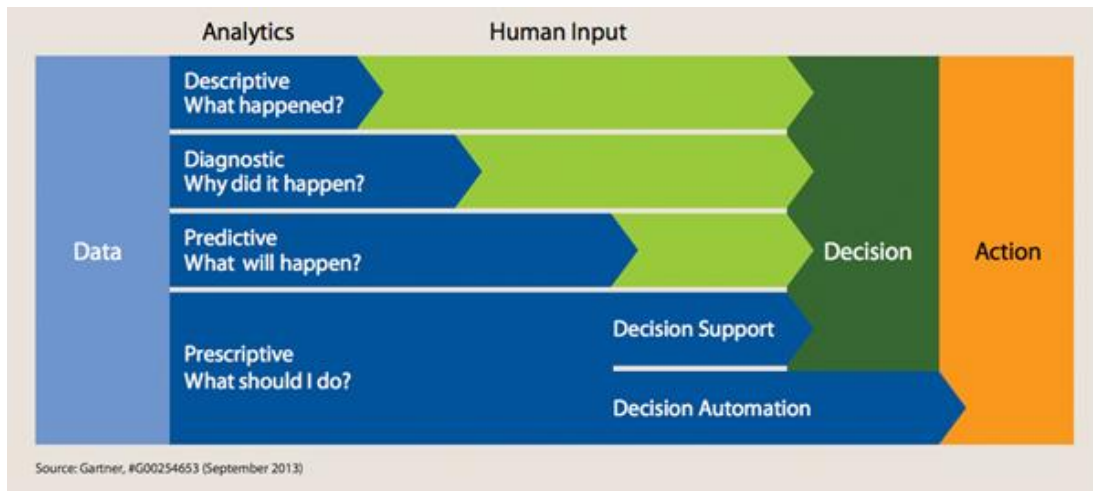
Figure 4: DAARC data analytics categories

The data analytics scope and assistance is based on:

- Obtaining access to appropriate governed and ungoverned data for analysis, discovery and research activities.

- Ability to join data from multiple sources and making business decisions by analyzing large volumes of data.

- Use Cases focused on innovation, knowledge discovery, using exploratory data analysis, prototyping, modeling, and systems infrastructure analysis.

- Users' awareness of authoritative data for querying various data warehouse subject areas and data marts.

- Specific piece of analysis, or peer review of analytics methodology.

- Extending from developing the appropriate research question to the full execution of the analytics project, and sharing actionable insights and improvement recommendations.

- Training and user enablement needs.

DATA ANALYTICS ACADEMY: USER TRAINING AND SUPPORT

The Data Analytics Academy provides customized classroom training, video-on-demand, webinars on Analytics techniques, methodologies, data coaching by agency data and business subject Matter Experts (SMES) data stewards, data owners custodians, and tool training on SAS® EBI and data analytics toolset. Enrolling in the Data Analytics Academy is a prerequisite, as it prepares the user for on-boarding into their respective data analytics projects and initiatives.

"Analytics Academy offers targeted, specialist analytics training across all components with the mission of improving capabilities across the agency"

Figure 5: DAARC Training Models

- The Data Analytics Academy is the educational component of DAARC

- DAARC extends analytics training courses to all interested components within the agency

- DAARC's mission is to teach analytics techniques designed to improve the accuracy, quality, and timeliness of services through data-driven decisions.

- Data Analytics Academy training courses include:

  o Data-driven techniques and methods

  o Predictive analytics

  o Text analytics

  o Numerical analysis

  o Anomaly detection

  o Rules engine

  o Data and Policy coaching

  o SAS$^{®}$ EBI Tools Training

  o Data analysis and analytics to improve operational efficiency

  o Using data to set expectations and improve staff productivity

  o Data visualization techniques

  o Natural language processing to analyze textual fields

PROMOTE DATA ANALYTICS REUSABLE METHODOLOGY AND ARTIFACTS

Stakeholders must answer business questions that can be sometimes repetitive and often require recreating information from previous efforts to meet requirements of a new data analytic request. One

of DAARC's main objectives is to make sure data analytics projects use analytics methods and processes that can be easily replicated or reused for other datasets and in different contexts

Creating Reusable Methodology/Artifacts for Advanced Analytics

Strengthening Analytics Techniques	Analytics Project Management	Socialization

- Develop reusable processes and tools to advance understanding of analytics techniques throughout the agency.

- Develop standardized templates to communicate understanding of advanced analytics artifacts.

- Develop project management methodology unique to advanced analytics projects.

- Develop standardized process for executing ad hoc advanced analytics requests.

- Integrate reusable artifacts into DAARC Communication and Interaction plan.

- Publish and socialize reusable artifacts with other analytics practitioners throughout the agency.

PROMOTE ADVANCED ANALYTICS COMMUNITY OF INTEREST

Today organizations, workgroups, teams, and individuals must work together in new ways. Inter-organizational collaboration is increasingly important. Communities of interest provide a new model for connecting people in the spirit of learning, knowledge sharing, and collaboration as well as individual, group, and organizational development. Spending resources and time helping to build such a community of interest will improve agency's commitment to building a culture of advanced analytics.
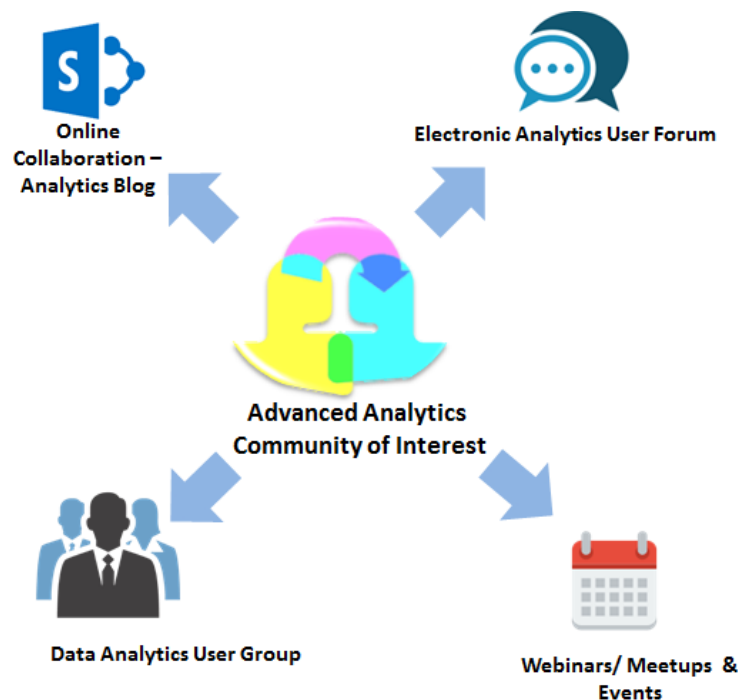
Figure 6: DAARC analytics community of interest

- *Connect people* who might not otherwise have the opportunity to interact
- *Enable dialogue* between people who come together to explore new possibilities, solve challenging problems and create new beneficial opportunities
- *Stimulate learning* by service as a vehicle for mentoring, coaching and communication
- *Capture and diffuse existing knowledge* to help people improve their practice by providing a forum to identify solutions to common problems and evaluate best practices
- *Generate knowledge* to help people mature their capabilities

ASSESS COMPONENT ANALYTICS NEEDS AND ASSIST IN INCREASING THEIR ANALYTICS MATURITY LEVEL

Analytics is defined as the scientific process of transforming data into insight for making better decisions, but what matters is the *purpose* of analytics. Analytics are used to gain actionable insights from data in order to make better decision. Various levels of analytics maturity can be distinguished, depending on how much of the decision process is automated, and how much is left for human intervention.

Analytics Maturity Assessment

- Components request DAARC assistance to assess analytics maturity
- DAARC performs an assessment against Industry Analytics Capability maturity model using it as a measure and guide
- The assessment will produce a roadmap to guide the component to achieve their desired level of analytics maturity.

Assistance to Improve Analytics Capabilities

- DAARC will provide direct assistance for standing up strong analytics units.
- DAARC will provide materials and recommendations for strengthening their analytics capabilities and capacity, and a culture of evidence-based decision-making.

DAARC Core Services Summary	Analytics Capability Maturity Model Attributes					
Core DAARC Services	Analytic Opportunities	Data	Analytic Techniques	People	Technology	Culture
Deliver Analytics Projects & Support Component Data Analytics Needs	✓	✓	✓	✓	✓	✓
Provide Data Analytics Training			✓	✓		✓
Deliver Reusable Analytics Methodology			✓	✓	✓	✓
Engage Analytics Community of Interest & Practice	✓		✓	✓	✓	✓
Assess Component Analytics Needs & Assist in increasing their Analytics Maturity level	✓	✓	✓	✓	✓	✓

Figure 7: Alignment of DAARC Core Services to the Analytics Maturity Model

AUTHORITATIVE DATA SOURCES AND ENTERPRISE METRICS CERTIFICATION

An authoritative data sources(ADS) is a cohesive set of data assets that provide trusted, timely, and secure information to support a business process.Information in ADS) is visible, accessible, understandable, and credible to information users.Enterprise metrics are used to evaluate factors crucial to the success of an organization. We use Enterprise metrics as a set of indicators to measure data against enterprise targeted goals. The agency's Enterprise metrics link to multiple business performance processes and agency performance goals. Since executives and managers use enterprise Enterprise metrics for both internal and external reporting purposes, it is of upmost importance that these metrics be reliable, and that users understand the authoritative data sources, business rules, and processes for establishing these metrics.
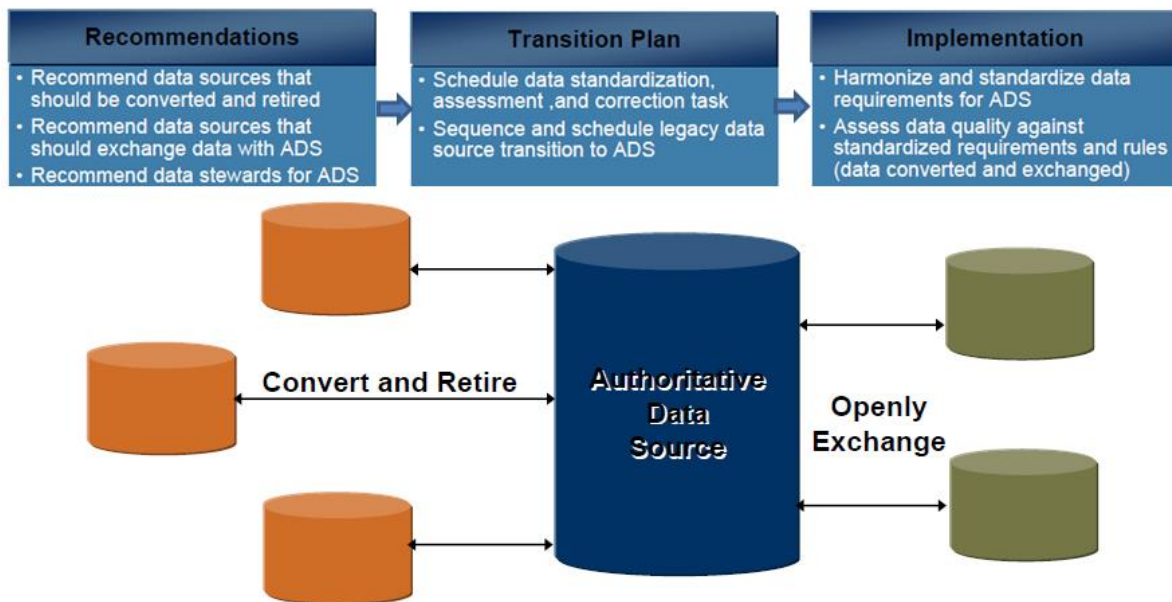

Figure 8: ADS Certification Process

The DAARC's responsibilities will include:

- Evaluating the agency's' Enterprise metrics; Identifying, documenting and publishing business process rules and data sources for calculating Key Performance Indicators (KPIs):
- Developing an iterative process to identify, analyze, test, certify, and document existing and new authoritative data sources used to support metrics associated with KPIs;
- Mitigating issues associated with disparate and conflicting KPIs and supporting metrics data across our systems;
- Recommend data sources that should be retired, exchange data and transition to ADS.
- Harmonize and standardize requirements for ADS.
- Functioning as the arbitrator between components when there is a difference in metric values, business rules application, or other issues;
- Identifying and making recommendations to eliminate redundant metric reporting; and
- Recommending that business rules associated with agency metrics align with our policies and procedures.

CONCLUSION

For any agency their leaderships' attention and support are critical, so make sure the analysis speaks to them; users will make or break the move to data-driven operations, so listen to them, make their work easier and make mission analytics a carrot, not just a stick; find ways to collaborate within and outside your organization to get data, analysis, expertise and even funding. DAARC provides a platform for these agencies to mature their data-driven programs for multiple audiences to dissect, discover and decide, hence turning data into knowledge. The main strength of DAARC is in reducing the data gatekeeper role in the organization so that every user can interact with the data to perform discovery and advance analytics sprints without being caught up report request backlogs.

Simply dumping reports, scorecards, data dumps, etc. on them does not work. Sending business managers to a small team of analysis gurus is nothing more than a bottleneck. Optimizing specific business processes with probabilistic models is not enough - the future of decision-making cannot become merely "machine driven. Data driven decisions run the nerve center of business, a data analytical and research center allows the business to make well-informed decisions while leveraging their talent's intuition and learning.

We hope this paper will help agencies to re-examine their data driven approach and consider implementing a DAARC for applying data analytics solutions to further their missions, save taxpayer dollars, improve services and more effectively achieve agency goals.

REFERENCES

- TDWI Best Practices Report - Integrating Hadoop into business intelligence and data warehousing by Philip Rossum: https://tdwi.org/Blogs/TDWI-Blog/List/Philips-Blog.aspx?Page=4

- The Obama Administration's Evidence-Based Social Policy Initiatives: An Overview by Jon Baron and Ron Haskins: http://www.brookings.edu/research/articles/2011/04/obama-social-policy-haskins

- Demystifying Business Intelligence, Business Analytics and Predictive Analytics by Gary Cokins http://blogs.sas.com/content/publishing/2013/11/21/demystifying-business-intelligence-business-analytics-and-predictive-analytics/

- From Data To Decisions I, II and III - Building an Analytics Culture, partnership for public service, IBM center for the business of government: http://www.businessofgovernment.org/sites/default/files/From%20Data%20to%20Decisions%20II.pdf

- TDWI Best Practices Report | Operationalizing and Embedding Analytics for Action, December 21, 2015: https://tdwi.org/research/2015/12/best-practices-report-operationalizing-and-embedding-analytics.aspx?tc=page0

- TDWI Best Practices Report | Next-Generation Analytics and Platforms, December 18, 2014 https://tdwi.org/research/2014/12/best-practices-report-next-generation-analytics-and-platforms.aspx?tc=page0

- How to use data analytics effectively in federal agencies by Tom Fox,   November 15, 2013: https://www.washingtonpost.com/news/on-leadership/wp/2013/11/15/how-to-use-data-analytics-effectively-in-federal-agencies/

- Optimize your federal workforce using data and analytics, Five Strategies for Federal Management to Maximize Human Capital : https://www.accenture.com/us-en/insight-optimize-federal-workforce-data-analytics.aspx

- Report: Agencies should use data to revitalize HR By Jake Williams, May 20, 2014 http://fedscoop.com/report-agencies-use-data-revitalize-gov-hr

ACKNOWLEDGMENTS

RECOMMENDED READING

- *Data Preparation for Analytics Using SAS by Gerhard Svolba, Ph.D.*

- *BIG DATA ANALYTICS through SAS by Cesar Perez*

- *Demystifying Business Intelligence, Business Analytics and Predictive Analytics by Gary Cokins*

CONTACT INFORMATION

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

Vivek Sethunatesan

Technical Advisor, Health IT

Northrop Grumman Information Systems, 2810 Lord Baltimore Drive

Baltimore, MD 21244

Email: vivekanandan.setunatesan@ngc.com

Linked in Link Vivek

Vivek Seth@BI360info