

PP12

Big Data Paradox: Development of a Unique and Repeatable Analytical Process while Implementing a Robust Data Science Solution

Introduction

Enormous amounts of data are being generated and are made available by the second. However, data has value only when it is utilized appropriately. It is very difficult to handle large data sets using traditional methods and techniques because the data may be big in volume, but also in variety, veracity, and velocity.

FAERS

We have developed a proof of concept using FAERS data provided by FDA to analyze and visualize signal detection and validate our methodology. What is FAERS?

“ The FDA Event Reporting System (FAERS) is a database that contains adverse event reports, medication error reports and product quality complaints resulting in adverse events that were submitted to FDA. The database is designed to support the FDA’s post-marketing safety surveillance program for drug and therapeutic biologic products.* ”

*FDA.gov

Data Science



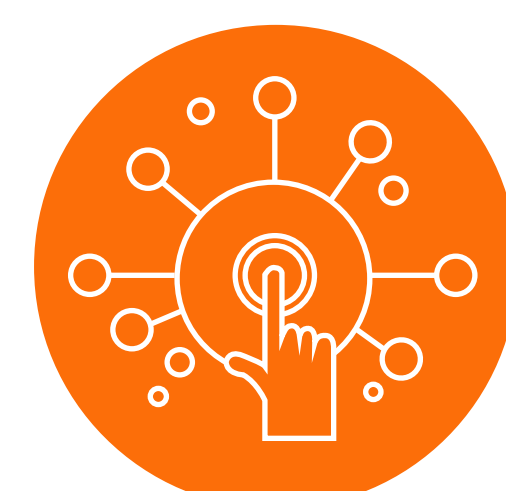
Discovery Identify where we can leverage data for making better business decisions, define goals and objectives of the analytic approach.



Data Engineering Gather data requirements, collect and build understanding of the data, clean and prepare data for analysis.



Algorithms & Analytics Analyze/refine the cleaned data to develop the model and analytics that pertain to the business problems.



Share & Decide Share valuable results and develop actionable insights with stakeholders, get feedback and repeat methodology as necessary.

Challenge

Apply a best-in-class data science methodology to analyze FAERS data for post-marketing safety surveillance and signal detection.

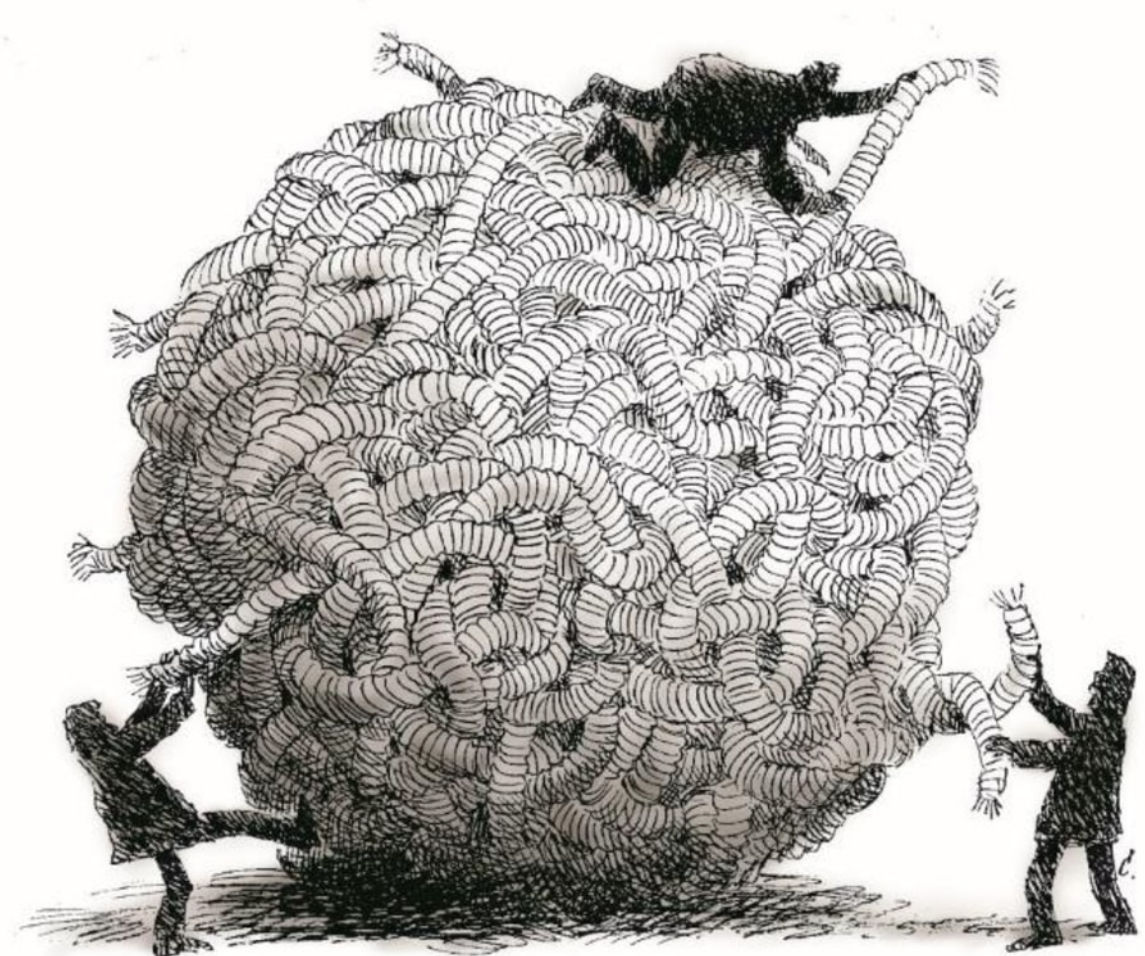
Conclusion

The proof of concept successfully demonstrated the methodology could:

- Produce a more efficient solution to meet the business need, in a relatively short time
- Quickly ingest and prepare analysis data from very large raw data sets very efficiently
- Apply an interactive, fast, and visually rich web-based analytics solution
- Allow us to elicit regular and in near real-time feedback
- Provide a consistent and repeatable process which can be applied to any Big Data project

Methodology

Discovery

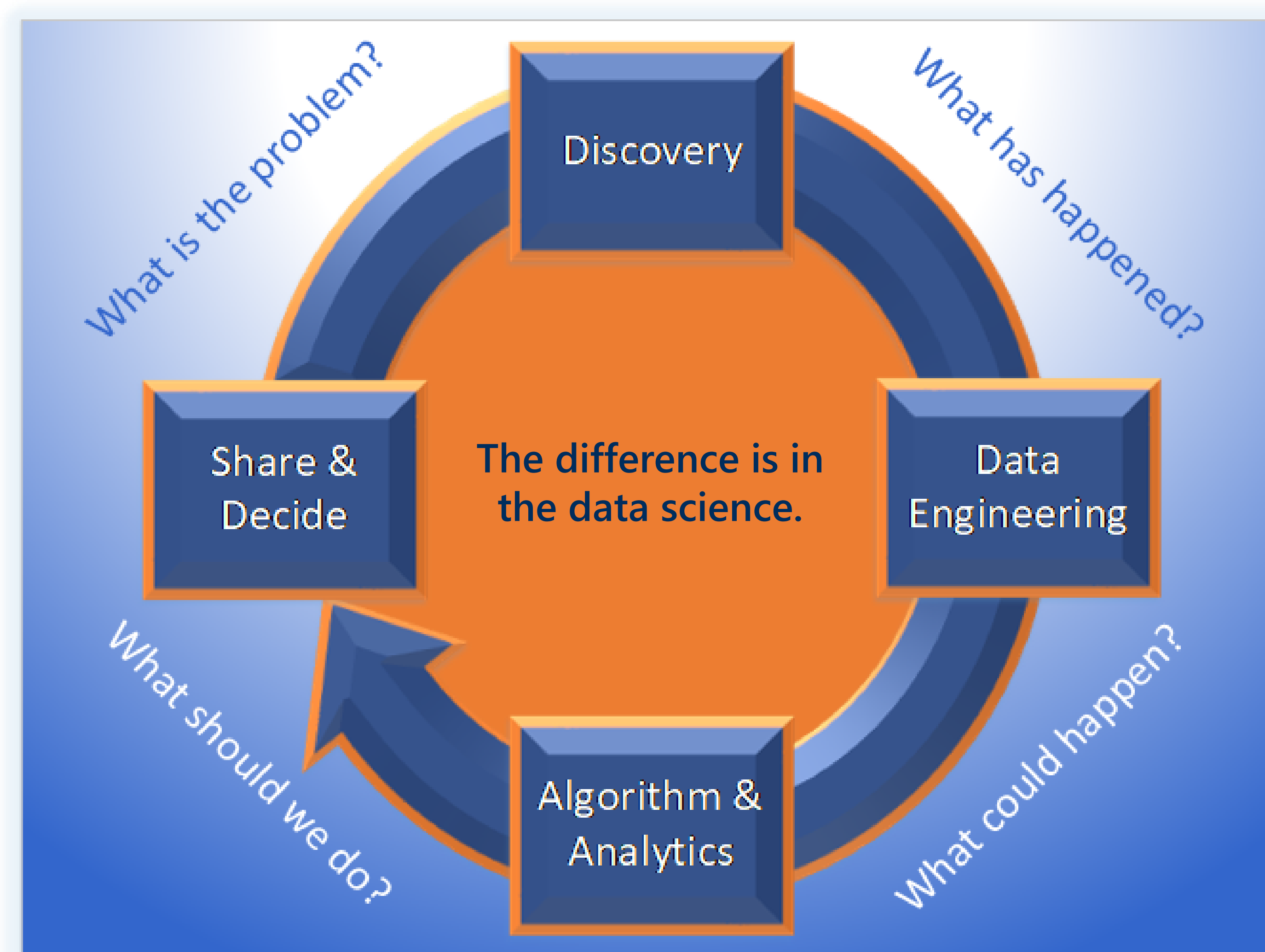
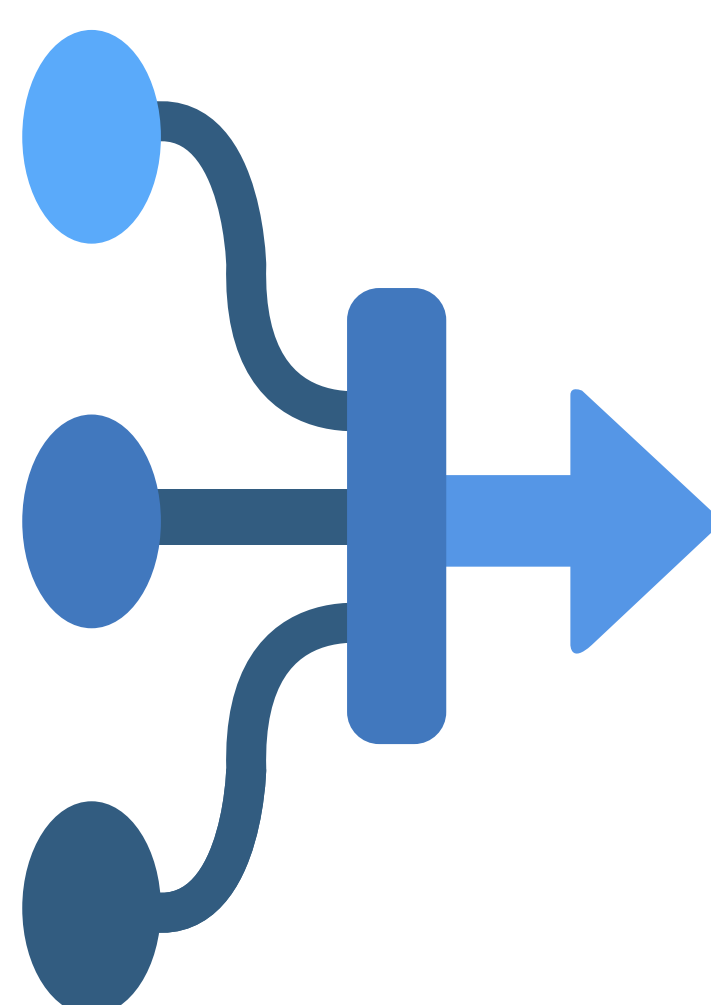


Data Engineering

Structured Data
Databases, Datawarehouses,
Enterprise System (CRM, ERP, etc.),
Census Records, etc.

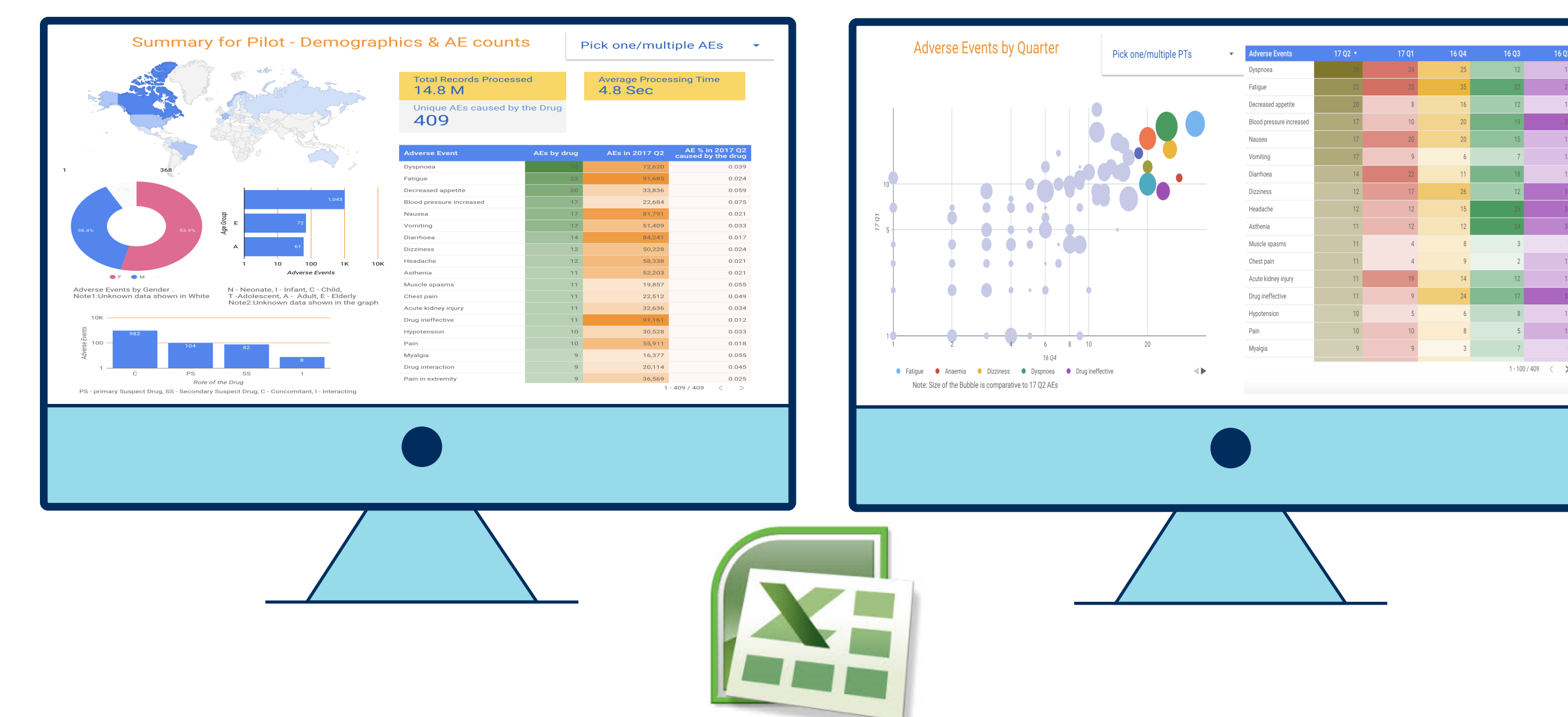
Semi-Structured Data
XML, E-Mail, etc.

Unstructured Data
Audio/video, X-Rays,
Audio/Visual, Natural Languages
in Health Records, etc.



Algorithms and Analytics

Share and Decide



Benefits of the Proof of Concept

- ✓ Validated MMS’s data sciences methodology
- ⚙ Leveraged publicly available and very large data volume to create value
- ⊕ Provided a quality solution for post-marketing surveillance

Chris Hurley
churley@mmsholdings.com

MMS
The difference is in the data.™