



Manufactured Simulated Training Data

Background

ExactData® is the world's leading provider of Manufactured Simulated Data Services. Current clients include some of the largest US Federal Agencies including the Department of Defense and Internal Revenue Service. ExactData's correlated, complex "smart" data is unmatched by any other company in the world. We manufacture our customized simulated data using a patented, sophisticated rules engine designed for the specifics of your requirements for realism, complexity, and scale. Our fully engineered smart data includes the unique features of longitudinal consistency, internal consistency, consistency across disparate data sets, and perfectly known ground truth. This enables comprehensive quantitative modeling, simulation and training against complex, realistic databases that have no privacy or confidentiality concerns.

Our unique Dynamic Data Generator™ technology doesn't require any access to production data or live records. Using this high-fidelity data, the training organization can complete complex modeling, simulation and training scenarios on demand. Many other Government Agencies have successfully used ExactData's Manufactured Simulated data: DARPA (for insider threat detection), U.S. Army (DoD Manufactured Artificial medical records), and IRS (Manufactured Artificial data for testing the additional complications of the Affordable Care Act on tax forms). ExactData has supplied Manufactured Simulated data to commercial enterprises also, such as IBM (Manufactured data for testing the telephony application in the 2010 Census), and more recently Oracle (Manufactured Simulated medical records for demonstrating their internally developed software).

Delivering On the Vision

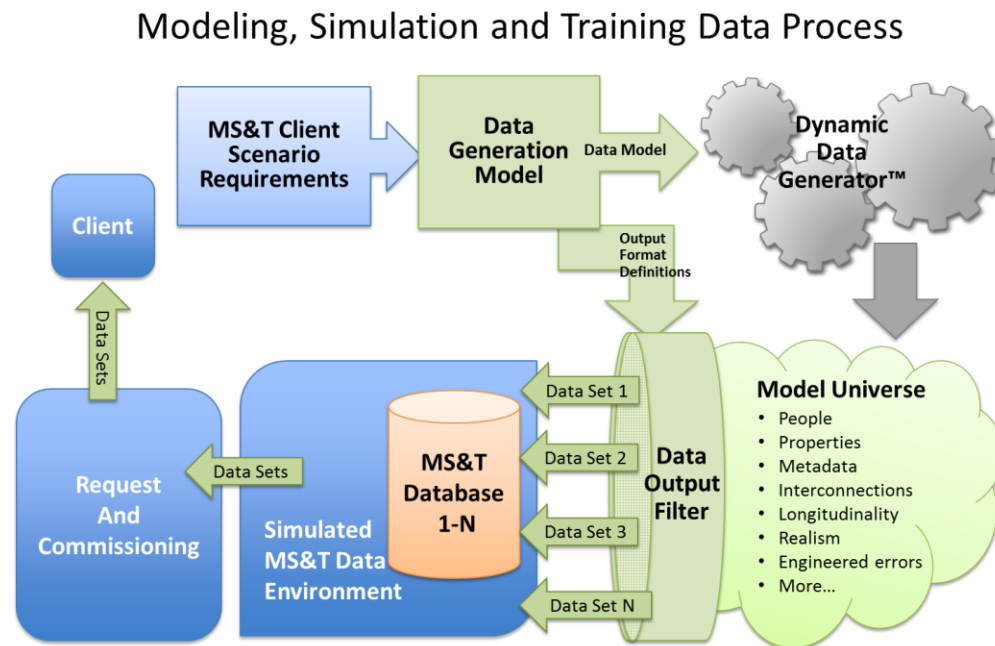
Manufactured artificial data technologies are mature and have been deployed on both large mission critical US Federal and commercial programs. The US Department of Defense is currently modernizing its Electronic Healthcare Records (EHR) system (DHMSM Project through Team Leidos), one of the world's largest EHR projects and expected to take over 10 years to complete and cost in excess of \$15B. The DHMSM Project Team Leidos has decided to rely on ExactData for the entire database of electronic healthcare records needed for training and simulation purposes. All training scenarios and simulation data will be generated on demand, without any risks of privacy or confidentiality issues. New complex medical scenario variations can be created almost immediately.

Solution

Training organizations have an opportunity to improve the current and future modeling, simulation and training (MS&T) projects by using this new technology to manufacture large volumes of data customized for their specific scenarios. This simulated data will be designed for the specific project and used with absolutely no confidentiality or privacy issues. This would enable the training organizations to efficiently run training and simulations accurately quantify human system error rates and performance while driving costs out of the current data management process. The training organization would be able to quickly and easily verify that training systems correctly identify human performance issues by using engineered errors, "dirty data", inconsistencies, and accompanying truth files that reference intended

scenarios for scoring. The manufactured simulated data also includes longitudinal consistency that correctly portrays historical events and genealogy over time.

The training organization could create sand box environments with simulated data systems and repositories. The sand box simulated data environments could be released to industry to solve simulation, modeling and training challenges and demonstrate effectiveness before purchase while increasing operational efficiencies in data management. This process enables more efficient development of tools and systems for simulation and training by decreasing timeliness of simulated scenario data from weeks and months to days and weeks.



1

Key Features

- Does not require access to an existing data source eliminating the risks of using production data
- Data attributes of realism, complexity, scale, internal consistency, consistency across disparate data sets, consistency over time
- Known ground truth files to reference for specific errors in data for the desired training scenarios
- Data can be produced virtually on-demand
- Produces large volumes of data and is easily scaled, as needed

Key Benefits

- Eliminate cost and risk of managing private/secured data
- Time Savings – access to data in days/weeks instead of weeks/months compresses IT development and implementation times while reducing cost
- Improved precision and accuracy - typically increased scenario case coverage from 15-30% to 60-80%
- Reducing labor cost through automation of large volume simulated data generation



- Saves the cost to correct training errors discovered post-deployment

Summary

Training organizations have many challenging goals providing innovative, inspiring MS&T work, while being good stewards of the organizations monies. Innovations in IT services will require creation of new infrastructure and tools. The use of manufactured simulated data repositories will improve security, cost effectiveness, availability and timeliness for implementation of the MS&T platforms.

For more information contact: Gail Tabor, VP and GM US Federal, (703) 407-6617, gail.tabor@exactdata.net