

Artificial Manufactured Data Repositories Enabling CDC Integration and Innovation

Background

The Centers for Disease Control and Prevention (CDC) faces a difficult task in the 21st century. The world is constantly changing and technology is exponentially increasing the number of systems that the CDC uses to respond to those changes. This creates a sea of interoperability issues, starting with improving the legacy systems at the CDC to work with the new tools, techniques, and information technology of the future CDC. A single system upgrade that supports the Meaningful Use (MU) for Electronic Health Records (EHR) can run into roadblocks with PII and HIPPA issues. With this potential issue for a multitude of systems interacting with health data repositories, the amount of time and effort to retire redundant surveillance systems while updating and developing new ones becomes extended; this costs more time and money that will reduce the availability and timeliness of data and its usefulness.

Delivering On the Vision

In order to improve the CDC Surveillance Platform, the CDC is faces a great many challenges concerning its proliferation of systems, the interoperability of silos, a slow adoption of new technologies, and developing with emerging health information policies. Configuring interoperability for messaging and interfacing for over 120 surveillance systems becomes very difficult. Meanwhile, the silos that these systems interact with contain inefficient interconnections and interdependencies. The lack of compatibility prevents deploying effective data analytics. This does not take into account that interfacing with many state and local health departments creates additional challenges with even more systems and more requirements. While some of the state and local health departments are keeping up to date with newer technologies for EHRs, the slow adoption of technology within the CDC hinders keeping up with these newer systems and the tools that are used for timely health surveillance. Lastly, EHR and MU standards create an even bigger roadblock for development, testing, and training with added interoperability requirements. This is where ExactData™ and its status as the world's leading provider of Manufactured Simulated Data Services can help overcome these challenges.

Using ExactData's Dynamic Data Generator™ (DDG), models can be configured within the DDG to simulate all of the 120+ CDC data systems to enable integration with newer systems and tools. The DDG manufactured artificial data can simulate the disparate data repositories to show interconnectedness in data while maintaining high levels of data integrity as well as embed errors and emulate incomplete or "dirty data", malformed data typical of the system to be tested. The added value of having simulated data environments also applies to any state or local health department system interoperability requirements. Maintaining high levels of data integrity for multiple systems across multiple data repositories can be a critical component when solving interoperability challenges and implementing new technologies. With ExactData's broad experience creating manufactured artificial EHRs, the CDC can efficiently and cost effectively implement emerging health information policies. This can all be done

while fulfilling interoperability requirements and eliminating the cost and risk of managing private and confidential information using the DDG.

Solution

The CDC has an opportunity to improve the current and future CDC Integrated Surveillance Platform deployments by using the patented, new DDG technology to manufacture large volumes of customized data. This simulated data will be designed for specific applications and uses with absolutely no confidentiality or privacy issues. This would enable the CDC to efficiently measure systems performance, accurately quantify system error rates, and accelerate algorithm improvements while driving costs out of the data management process. The CDC will be able to test and deploy analytics more effectively and with increased accuracy. The CDC will be able to quickly and easily verify that systems correctly identify interconnectedness and analytic issues by using engineered errors, "dirty data", inconsistencies, and accompanying truth files that reference intended scenarios in the for scoring. The manufactured data includes longitudinal consistency that correctly portrays historical events and genealogy over time.

To address interface issues, the DDG would create data for sand box environments, emulating CDC data systems and repositories. These environments can exist on or adjacent to the CDC Surveillance Platform allowing for specific application and system needs. The sand box simulated data environments could be released to industry to solve CDC challenges and demonstrate effectiveness before purchase while increasing operational efficiencies in data management. ExactData's process enables more efficient development of tools and systems to test interoperability by decreasing timeliness of data from weeks and months to days and weeks.

Working closely with clients to define the artificial data required, the data is built to specific system requirements for realism, complexity, and scale with absolutely no confidentiality or privacy issues.

ExactData delivers positive results for your projects

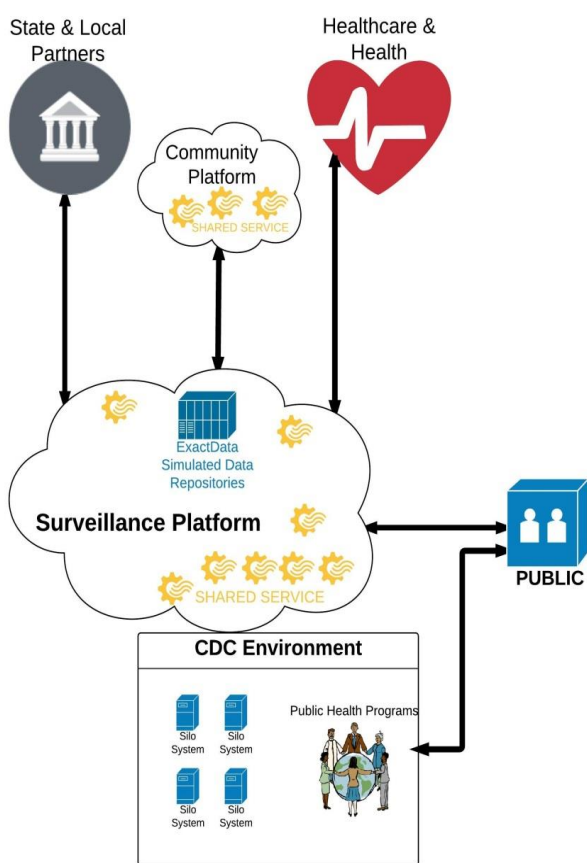
- Reduce the cost of test data creation by 90%. A typical agency has over 500 full time equivalents involved in this process yielding savings of \$50M per year.
- Reduce project development timelines by up to 50%. This directly translates to cost savings for measured in \$100M's per year.
- Increase use case coverage and measure error rates in your development environment. Correcting errors in the development process and reducing the cost of escape errors will save your agency \$100M a year.
- Eliminate the cost and risk of managing confidential and private information in your development environments.
- Create collaborative Industry environments where challenge databases are released to the public and Industry and prototype and prove value before you purchase.

Benefits to the CDC

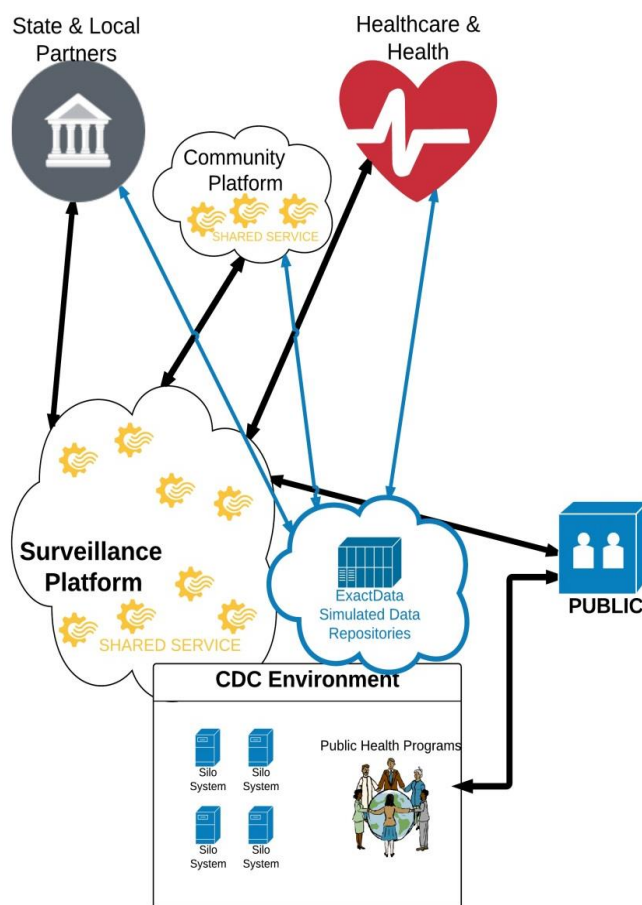
- Improve predictive analytics and threat detection, reduce fraud

- Require NO access to an existing data source eliminating the risks of using production data; eliminates data security risk for development databases and a cyber threat vector
- Reduce the cost to correct errors in production systems missed in the development process
- Rapidly manufactured artificial test data reduces overall development schedules and costs and enables Agile development
- Inter-agency collaboration during the development process doesn't require any access to sensitive data
- Rapidly produces large volumes datasets

Common Current Platform



Expanded Capabilities with Artificial Data



Summary

The CDC has many challenging goals playing an integral role formulating how health data and public health systems are used for surveillance purposes. CDC innovations in health IT services will require creation of new infrastructure and tools. The adaptability of deploying the repositories in whatever fashion will aid the interoperability in all its facets. The use of manufactured simulated data repositories, will improve availability and timeliness for implementation of the CDC's surveillance platforms.

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