Sanofi Accelerated the Standardization of Clinical Trial, Marketing and Commercial Data to Deliver New Insights on Consumer Health and Drug Development

Company Background
Sanofi is a global healthcare company engaged in research, development, manufacturing and marketing of therapeutic solutions. Sanofi has core strengths in diabetes solutions, human vaccines, innovative drugs, consumer healthcare, emerging markets, animal health and Genzyme. Sanofi was founded in 1973, with a long history of chemistry and manufacturing.

With more than 110,000 employees in 100 countries, Sanofi is focused on becoming a trusted part of every individual's life. “Rather than think of individuals as patients, we think of them as healthcare consumers because people want to live their lives as normally as they can and you’re really only a patient when you are seeing a doctor or you’re in a hospital.” says Brian Ellerman, Head of Technology Scouting and Innovation.

Data Challenges
Sanofi faces the challenge of having multiple entities within its portfolio that are all stand alone companies utilizing their own data sets to drive decisions. Sanofi’s data management goal is more seamless interchange data between entities, moving from silos to enabling transparency across business units and ultimately towards a more universal health care consumer vision.

Translational Medicine is an important area of research at Sanofi—it is an area of biomedical and public health research that aims to convert key discoveries made in research labs into drugs, tools and medicines for patients. Sanofi’s significant challenge resides in cleaning biomarker and clinical data, which is extremely messy and noisy (from data inputted incorrectly or corrupted in a processing step); to produce meaningful information for genomics and molecular biology teams to draw correlations between the biomarker and clinical outcomes. In cancer therapy, there are a number of biomarkers that can be used to predict whether or not a person is going to be responsive to a certain chemotherapy agent. If the data is noisy at the biomarker stage, noisy in the research process and noisy in the clinical process, it would be difficult to predict how a cancer might present itself. By cleaning and transforming the data, Sanofi is able to uncover the valuable signal from the noise.

Prior to working with Trifacta, Sanofi addressed these data management needs by utilizing individuals with deep medical backgrounds as well as very strong biostatistics skills, to manually curate data. This resulted in the challenging scenario of having to use their low supply of highly skilled staff, with its high cost structure, to spend most of all of their time cleaning data.

“...number one, it doesn’t require a highly skilled data analyst to operate the software and we can really put this in the hands of anybody. Automating the data curation aspect has a massive benefit because I can get the data scientists to work on the really hard stuff instead of working on the data cleansing. Number two, I can put in the hands of the average front line business person, something that can let them do a lot of the ‘what if’ modeling without having to ask an overburdened resource like a data scientist. The rapid implementation of course makes it all the better. To me, besides the user friendliness, the one resounding positive that I heard again and again was how quickly people were able to go from, I’ve got this amount of data to I’m actually doing something with it.”

MR. BRIAN ELLERMen
Head of Scouting and Innovation, Sanofi
Sanofi validated Trifacta’s data wrangling usability, scope and scale in the preparation of a variety of biopharmaceutical data sets. With Trifacta’s approach to data wrangling, users of varying technical abilities can discover, structure, clean, enrich, validate and publish data for analysis.

**Trifacta Solution**

Sanofi leverages Trifacta for numerous types of data in different settings around the world enabling various business functions:

- Sanofi’s Translational Medicine for Patients team uses clinical and biomarker data to unify and standardize data from multiple clinical trials to provide R&D insights.
- The Marketing group processes commercial data inputs from European countries in order to gain insight and include local data in global analyses.
- The Medical Affairs Group produces field CRM reports related to Medical Inquiry Response (MIR). Keyword searches across activity descriptions provide customer insight from direct field engagement.
- Sanofi’s IT Innovation lab combines commercial, consumer health (CHC) heterogeneous data sets on over the counter products in order to improve their knowledge of customer profile and correlate events with client behavior.

The project teams were interested in finding a data wrangling solution that was easy to on-board, user-friendly for technical and non-technical users and that significantly accelerated the process of curating and cleansing disparate data for insights and analysis. Project team members were selected based on their backgrounds in science, rather than being skilled in information technology or having a deep data analytics background.

**Results**

Trifacta has enabled Sanofi’s researchers, scientists and data analysts to structure and curate numerous complex data sets from a variety of regions, internationally. Sanofi endorsed the project’s success based on Trifacta’s ability to execute on these criterias:

- Easy onboarding
- User-friendliness
- Rapid implementation
- High accuracy in operations
- Repeatability
- Support of JSON, CSV and Excel
- Timely product support services
- Time savings in the data curation process

Trifacta succeeded in all of these validations and, in particular, helped in shrinking the curation process from 2 weeks with their existing tools, to 1 hour using Trifacta.

At Sanofi, return on investment is not easy to measure in hard dollars. It’s often calculated in soft dollars related to patient life and quality of life measurement. Sanofi has developed an industrial and partner network strategy to support its growth platforms and foster innovation for patients’ needs. Trifacta’s data wrangling solutions helped them rapidly accelerate the data preparation process that will lead to new insights in drug development for a myriad of healthcare customers around the world.