



Case Study

Breaking Silos to Enable Breakthroughs

By Adam Kraut, Bill Van Etten, and John Jacquay - BioTeam, Inc.

For most big pharma and other global enterprises, decades of growth have resulted in siloed operations, even as they seek breakthrough medicines, particularly in the wake of the current pandemic. The 2003 discovery of the biomarker PD-L1, Programmed Death Ligand-1, a transmembrane protein associated with immunosuppression and implicated in numerous diseases, led to the development of a new class of anti-cancer drugs and shepherded in the advent of immuno-oncology, leveraging a person's own immune system to attack cancer cells. The discovery of this biomarker was one of the largest breakthroughs in the field of bioscience in recent history. It is time for another one; as an industry, we're overdue.

Given the unprecedented levels of data being generated, advances in technology, and improvements in analytical capabilities, it is somewhat surprising that there hasn't been another game-changing biomarker discovered.

Bristol-Myers Squibb (BMS)'s Informatics and Predictive Sciences Business Partner, Daniel Huston, believes that BMS may be sitting on the next breakthrough biomarker or blockbuster treatment, but the issues with data findability, accessibility, interoperability, and reusability make it a challenge to unlock the insights that would lead to such discoveries.

BMS wanted a data science solution where they could bring their questions to their growing wealth of their data. BMS also wanted to reliably compare results across multiple data domains and to have a systematic approach that offered greater reusability so that they could maximize the ROI of that data. Breakthrough discovery was going to require a whole new way of thinking, and a new solution.

Like most enterprise companies established decades ago with a history of growth that includes both organic expansion plus M&A, BMS was operating with an unknown and unmeasured number of disparate technology solutions—few of which were interoperable. Internal silos have long been recognized as an impediment to achieving corporate goals. Silos hamper internal communication and can promote team dynamics where territorial challenges thwart collaboration.

Box 1 - Examples of operational silos

Data Silos: data can neither be found nor accessed, hence it is not interoperable and not reproducible.

Application Silos: legacy, proprietary, commercial and open-source applications are often used by only one researcher or one department and may be hosted on a local system versus part of a centralized, secure, and managed network.

Human Silos: the flow of information becomes restricted as a byproduct of cost-center assignments, team dynamics, or other barriers.



“The question we asked ourselves was, how do we seize on the unprecedented opportunities that advances in data science, hardware, and theory offer to unlock transformative insights from our ever-growing wealth of data?”

Daniel Huston, Informatics and Predictive Sciences Business Partner, Bristol-Myers Squibb

Perhaps even more importantly, data silos prevent people from finding or knowing which data assets they have at their disposal. How do you request access to something that you don't know you have?

The data assets of a big pharma, including legacy data, could hold the keys to a breakthrough waiting to happen—but, without breaking down the silos, the likelihood of a breakthrough discovery is greatly diminished. BMS' research workflow had come to be standardized as follows: ask a question, generate data, analyze and interpret it, archive it then move on to the next question.

This workflow created multiple types of silos (Box 1) and exacerbating the problem of silos was the rapid growth experienced by BMS around two years ago when their bioinformatics group doubled in size. Newly onboarded scientists were unable to easily find or access the data and tools they needed; once they secured access to the data, they were unable to make sense of it. BMS realized that they needed to look outside for a solution to their data management challenges.

Following a comprehensive internal capabilities assessment effort conducted with the support of the BioTeam, BMS assessed the scope of their data problem, identify and prioritized their needs, then drove consensus around a plan of action. BMS quickly realized that further updates to a legacy system of silos was not going to generate the breakthroughs that they were seeking. What they needed was a new and novel system that could be customized to meet their unique needs and common challenges of today and the foreseeable future. The options they considered—but rejected—included:

- **Building a novel system from scratch (too labor-intensive and time-consuming);**
- **Purchase a commercial solution then customize it (too expensive and would require multiple stitched together solutions, further exacerbating a proliferation of tools);**
- **Leverage an open-source solution as a foundation that could be enhanced (risky to rely on a solution at the mercy of the public domain).**

An extensive analysis was conducted to weigh the benefits and limitations of each option; the analysis identified a set of non-trivial requirements. The solution

Box 2 - Key Capabilities of the Genomics Discovery Hub

1. Search and find datasets that span divergent systems;
2. Enable single sign-on with universal access, authentication, and authorization controls;
3. Provide graphical workspaces and customizable dashboards for non-programmers;
4. Integrate with other systems, tools, and processes via open APIs;
5. Monitor logins with audit-ready capabilities to meet the security and compliance standards for protected healthcare information and data privacy;
6. Import, tag, and index over 5 petabytes of existing data;
7. Curate an application catalog for tools to analyze the data;
8. Facilitate cross-study analysis with sharing capabilities.

would have to be installed on BMS' system and be compatible with existing infrastructure, security protocols and regulatory requirements. Activation energy to get it running could not detract internal efforts from ongoing research. It had to scale to meet the enormous volumes of data. And it could not rely on armies of IT staff to maintain it.

The first task was to clarify the collective understanding of the scientific and technical requirements then to drive consensus across all stakeholders around the top priorities. Together, the BioTeam and BMS identified eight capabilities (Box 2) and developed a plan to address each need individually and as part of a total solution.

Gen3, an open-source data commons platform, was selected because it could be tailored in a timely and cost-effective manner, as well as maintained at a reasonable cost and effort. Gen3 became the foundation for BMS' translational bioinformatics solution: The Genomics Discovery Hub, or GDH, which was designed to "free

their data.” The platform was assembled on the AWS Cloud in a stepwise manner (Box 3)

The GDH automatically captures and indexes data at the source, leaving it where it is, then brings compute environments to the data itself. This protects the original source data from inadvertent manipulation errors and saves costs from having to have multiple copies of a single dataset. Exploration and query interfaces allow for finding and analyzing datasets. A graph data model (see Figure 1) allows for standardization of metadata elements across the entire ecosystem, enabling interoperability and reusability of data. Anyone with access to the GDH, can now readily gain insights from the data and conduct their own analyses, which is a level of collaborative discovery previously not possible.

Collaborative discovery has now been undeniably enabled across the entire BMS enterprise. Today, more than five petabytes of cleaned, enriched, and searchable data are available to hundreds of scientists who no longer need to rely on programmers, analysts, statisticians, and IT experts to locate and analyze their data. The tailored and intuitive dashboards (Figure 2), along with graphical search capabilities, are now part of the standardized tool set available to all researchers and external collaborators. Because the data can be readily partitioned, tracked and

Box 3 - Steps used to create the GDH on AWS

1. Install the Gen3 platform;
2. Index and enrich over 5 PBs of legacy data;
3. Create data dictionaries and harmonization rules;
4. Automate the data loading processes;
5. Modify and build new microservices;
6. Build a service catalog.

associated, BMS now enjoys new efficiencies related to in- and outlicensing arrangements, M&A, and other collaborative efforts.

BMS has already seen a remarkable improvement in data understanding and the speed of analysis. It is now readily possible to quickly identify a variant (or collection of variants) or biomarker that holds the potential as a therapeutic candidate. They already recognize how their

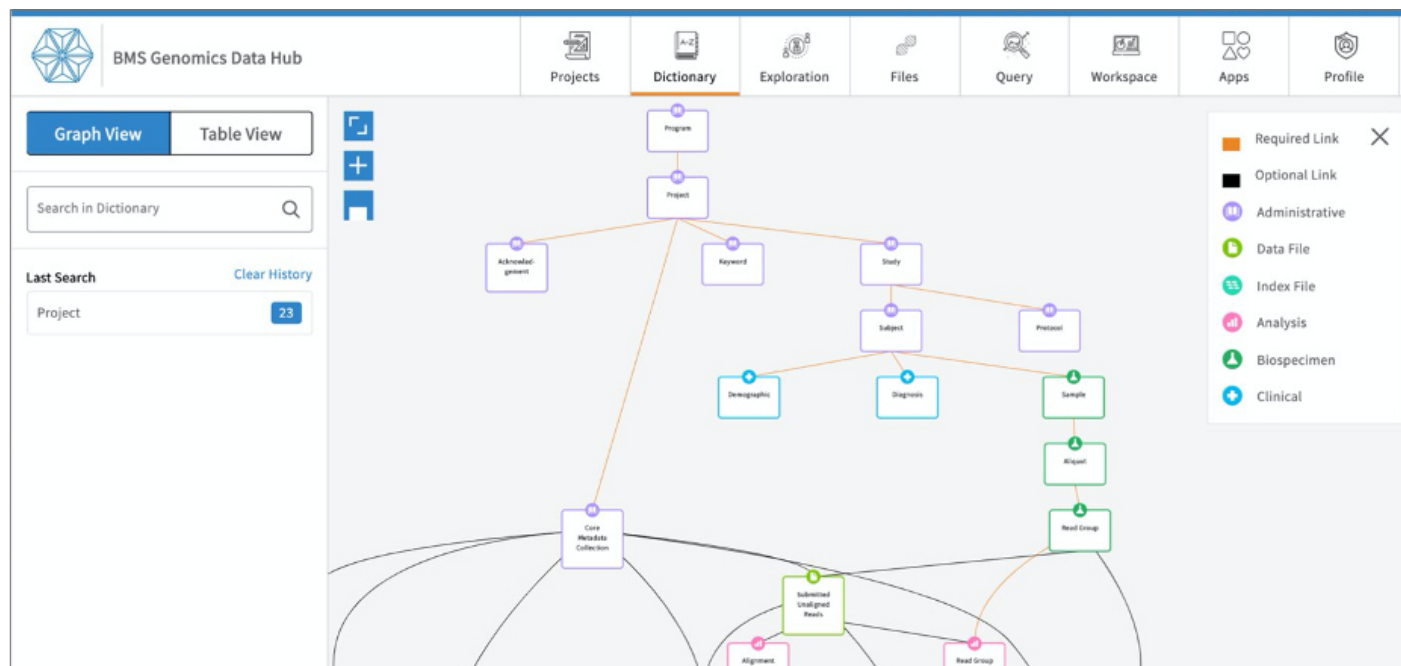


Figure 1 - Graphical data modeling with a Gen3 data dictionary

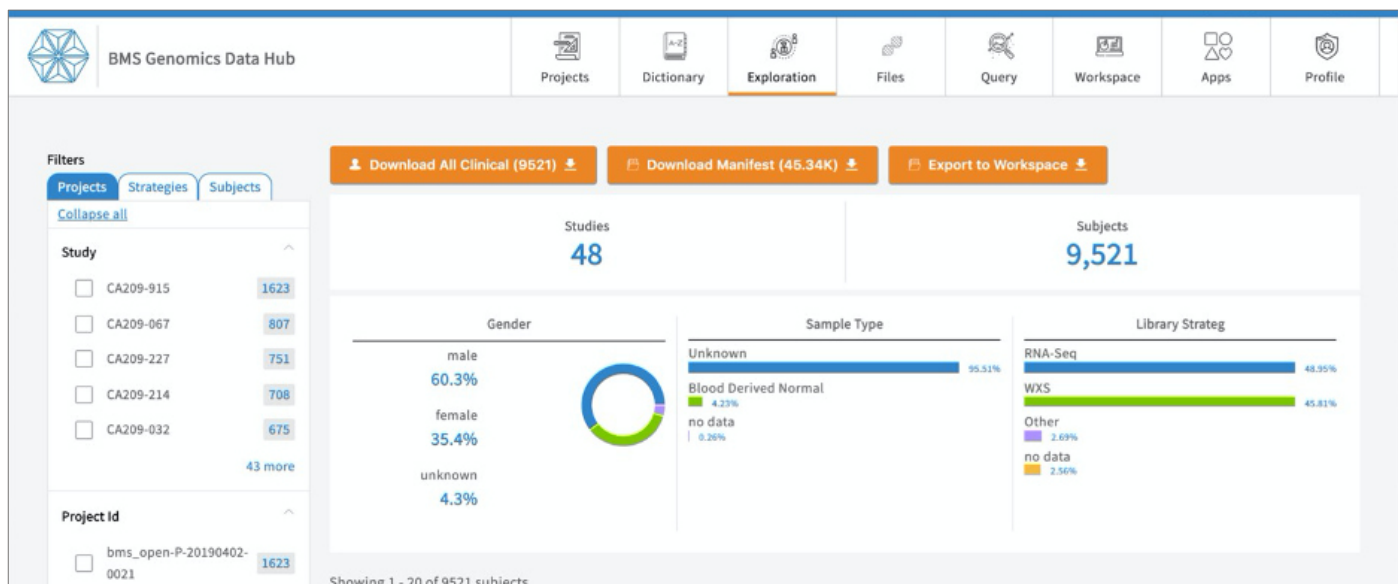


Figure 2 -: Faceted search and data exploration across the Gen3 data commons

solution will accelerate FDA approval and hence, their ability to bring life-saving drugs to the market sooner. Breaking their silos has been internally hailed as a game-changer.

Next steps will include extending the system beyond Genomics and including additional data domains such as Flow Cytometry, Immunohistochemistry, and Proteomics. We are no longer thinking about building a data commons for genomics, but rather increasing the productivity of all research and early discovery at BMS with a holistic and evolving data ecosystem. We think this will enhance BMS' ability to garner new insights from the data that they already have, independently, or in

association with the new data they are creating today and will create in the future. "At this pace, we may not have to wait another 20 years for the next breakthrough," affirms Huston regarding the success of the approach thus far.

For more information on creating data ecosystems like the BMS Genomics Discovery Hub, please contact BioTeam at:

info@bioteam.net

biote.am/breakingsilos