

COLLABORATIVE RESEARCH IN THE LONG-TAIL OF COVID

TAKEAWAY

Palantir has found five challenges that every modernizing research organization will have to address in order to remain competitive. These challenges emerge as the result of long-term trends familiar to anyone in the space.

Amidst the success of multiple COVID-19 vaccine development programs, it's easy to adopt an optimistic view of the future of pharmaceutical research. Indeed, there is plenty to be excited about, including AI and ML automating repetitive research tasks, real-world evidence (RWE) providing a tighter feedback loop on post-market drug data, and IoT/wearables increasing the amount of data available for analysis. At the same time, each new breakthrough is more difficult to achieve than the last, as reflected in declining overall returns on investment in pharmaceutical R&D. A 2018 study of the top 12 pharmaceutical companies found that since 2010, the average cost to bring a drug to market had doubled to $2.1 billion, bringing the total return to less than $1.02 for every dollar spent (Deloitte).

At Palantir, we work with some of the world's premier research institutions including pharmaceutical companies, health regulators, public health services, and research hospitals. This unique vantage has allowed us to identify salient differences and similarities between organizations.

At its most essential, a research institution, whether academic or commercial, transforms individual expertise into institutional knowledge. Customers tend to agree that people are their most important investment that can be scaled with the computational power of technology. In particular, regardless of the AI bluster, humans are critical to the novel scientific insights process. In light of this, it's clear why digital transformations - many of which aim to take humans out of the loop - fail so often. Palantir has always sought out human-computer symbiosis and has spent nearly two decades developing and deploying software that facilitates it. We believe this idea will be foundational to the research institutions of the future and today's technology investments should be directed toward this outcome.

A symbiotic approach to technology is exceptionally difficult to implement at scale. It entails organizing the experts already at these organizations, exposing them to the data, tools, and collaborators they need to quickly generate and validate ideas, and ensuring that all the research outputs - final results, intermediate data, useful pieces of logic, or subtle suspicions - become useful pieces of the institution's knowledge base. The specific path to symbiosis is unique for each organization, but the overall shape of it is not. Palantir has met customers at various points along this journey, offering us a chance to spot patterns that a one-time traveler would miss.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

From this experience, we enumerate five challenges that every modernizing research organization will have to address in order to remain competitive. These challenges emerge as the result of long-term trends familiar to anyone in the space: data sets are getting bigger in order to detect more subtle effects; data types are getting more complex with the proliferation of imaging, -omics, and IoT sensors; and privacy and security concerns are becoming more central to patients, researchers, and regulators.

These changes have been celebrated by the research community for good reason, and the organizations who can effectively address the emergent challenges will be situated to participate in what we see as the future of research, outlined at the end of this document.

DEEPENING SPECIALIZATION AND NARROWING DOMAINS OF EXPERTISE

TAKEAWAY

New insights are often expedited with cross-functional teams and multi-disciplinary team members who understand the world through different lenses.

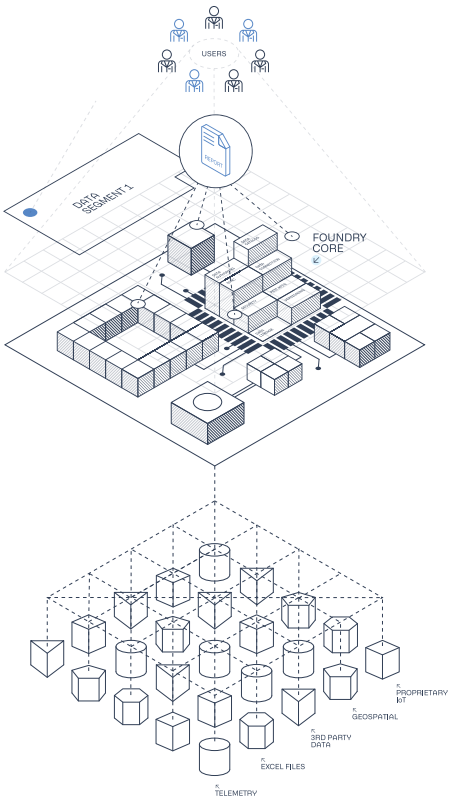
Modern research requires collaboration between several specialties, but these specialties grow more distinct every day. Within the field of oncology, for example, there are hundreds of types of cancer, each with their own robust and ever-deepening field of study. Even within closely adjacent fields like these, researchers struggle to find collaborators with complementary skillsets and, once found, they struggle to communicate with one another. Their knowledge is so specific and voluminous that bringing on a new collaborator takes years to decades of training.

While this problem is a natural consequence of the expansion of scientific knowledge, it is a serious impediment to progress nonetheless. New insights benefit from not just depth of expertise in a hyper-specific field, but also ideas, techniques, and equally specialized knowledge from fields both near and far from the problem. New insights are often expedited with cross-functional teams and multi-disciplinary team members who understand the world through different lenses.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

01

How Palantir's Foundry Mitigrates the Fracturing of Expertise →



Flexible Ontology Shifts With Understanding

→ The goal of scientific research is to incrementally construct a more accurate understanding of the world. Software systems tend to struggle with a constantly mutating data model, but Foundry was built to enable organizational adaptability. The Foundry ontology can be mapped and re-mapped countless times to ensure that it never becomes a relic of how your organization understood the world of the last decade or even last week.

Multiple Semantic Layers Allow Everyone to Understand the Data

→ Some experts, like data scientists, are used to thinking about the world in terms of tables, rows, and columns. Others prefer to think in terms of patients and diagnoses, or compounds and reactions. Ultimately, all of these languages are describing the same world. Foundry accounts for this by allowing users to build multiple semantic layers on top of the same data and interact with data through their preferred semantics.

No-Code, Low-Code, and Code-Based Tools Allow Everyone to Contribute

→ A research platform that requires coding proficiency will exclude much of the world's most important scientific experts. This is bad for those experts, but even worse for the institution who depends on them. Foundry provides analytical tools that span the gamut from point-and-click user interfaces all the way to high-scale data transformation code. There are multiple low-code tools available that accelerate research workflows without sacrificing flexibility.

Work/Collaborator Discovery

→ Over time, organizations accumulate more expertise and tools, but they also grow in scale and complexity. Finding the right tools, people, and prior learnings to accelerate your work can be extremely difficult. As Foundry pieces together a complete picture of who is working on what, it facilitates collaborator discovery to form more impactful research teams, and matches those teams with tools that accelerate their work.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

DEPENDENCE ON SPECIALIZED AUXILIARY FUNCTIONS

Scientists rely on a variety of auxiliary functions, that are each becoming more specialized every day. A 2001 paper sharing the initial sequencing of the entire human genome boasts nearly 3,000 authors listed. The 2012 paper sharing the discovery of the Higgs boson particle has around 5,000 authors. These massive numbers, whether they be overly liberal with attribution or excluding an even longer list of unnamed contributors, reveal an underlying fact: big results increasingly require big teams. This mirrors our own experience building the NIH's National COVID Cohort Collaborative (N3C), whose publications sometimes list 45 authors from 29 institutions. In addition to the listed authors, there is a small army of people behind the scenes doing the digital equivalent of organizing, moving, locking, and unlocking filing cabinets. In N3C's case, those 9.1 billion rows of patient data would require the organization of nearly 4.5 billion sheets of paper (LexisNexis, assuming Microsoft Excel files).

At high data scale, even seemingly basic tasks require specialized knowledge and tools, many of which are beyond the grasp of scientists. Thankfully, these problems are the sole focus of entire fields of study and maturing areas of expertise: data science, data governance, distributed computing, and semantics, to name a few. A technology strategy that fails to recognize the need for coordination and discovery among several distinct disciplines will exacerbate rather than alleviate pressure at the interfaces between them.

02

How Palantir's Foundry Synchronizes Multi-Functional Research Efforts →

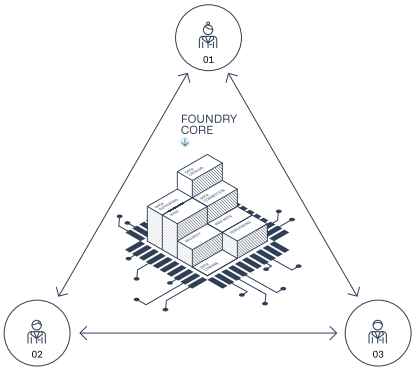
Productization of Expertise

→ A common pattern we see among our customers is a severe bottleneck on data science capacity. In addition to reducing cycle times by creating a shared collaboration space for domain experts and data scientists, Foundry allows people on both ends of the spectrum to productize their expertise. Data scientists can code common visualizations and statistical tests once and provide them thousands of times to domain experts through point-and-click interfaces. Likewise, domain experts can encode information like phenotype definitions into snippets that can be reused by data scientists in their studies.

Attribution That Incentivizes Collaboration

→ Automation can help scale users to achieve greater research throughput and Foundry can assist with ensuring attribution for those outputs. Foundry tracks who built each reusable artifact and who uses it, offering a way to recognize the people who scale up their impact by supporting their colleagues. With assurances that their multiplicative impact is recognized even if automated, experts can spend their time on the problems that demand their own personal touch.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.



Programmable Data Governance

→ Every organization has its own process to evaluate proper data access, whether it be an internal review board, an external data access committee, or by requiring authorizations from several data owners across the organization. Foundry's programmable data governance encodes this into the platform's core behaviors so that data providers, controllers, and consumers have a clear understanding of what data is being used where.

COLLABORATION ACROSS SPACE AND TIME

Research requires more than effective collaboration among colleagues in the same lab. To maximize returns on research activities, researchers must synthesize the work of prior generations of scientists and be responsible collaborators with their unknown colleagues of the future. This is traditionally done through a publication process, internally for most commercial organizations or externally for most academic ones. In either case, the overhead involved can often be high enough to become a deterrent to sharing, meaning the vast majority of research that an institution pays for is completely lost.

TAKEAWAY

Research technologies must dramatically reduce the cost of publishing - internally or externally - while also accounting for an explosion in data scale, diversity, and sensitivity.

In academia, article reviews regularly take 3-6 months, non-native English speakers face impediments unrelated to the quality of their science, and negative results are rarely published. Even these problems are only encountered after the largest barrier has been overcome: the researcher deciding it's worth even trying to share their work. The reality is that it's too expensive, in time and money, for scientists to attempt to share everything they find – even if they are confident it would be valuable to someone else.

In commercial research, the problem is similar but even harder to detect. Instead of knowledge being shared through curated journals, "publication" takes the form of email threads, meetings, unstructured documents, and water cooler conversations. Even at top tier pharmaceutical companies with whom we partner, well-meaning scientists frequently opt against doing extra work to catalogue their findings. This is a perfectly rational response to high personal cost and no personal benefit for sharing negative, duplicative, tangential, low-confidence, or what may be received as "uninteresting" results. Research technologies must dramatically reduce the cost of publishing - internally or externally - while also accounting for an explosion in data scale, diversity, and sensitivity that introduces technical challenges to sharing.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

03

How Palantir's Foundry Helps Your Organization Share Knowledge →

Zero-Cost Internal Publishing

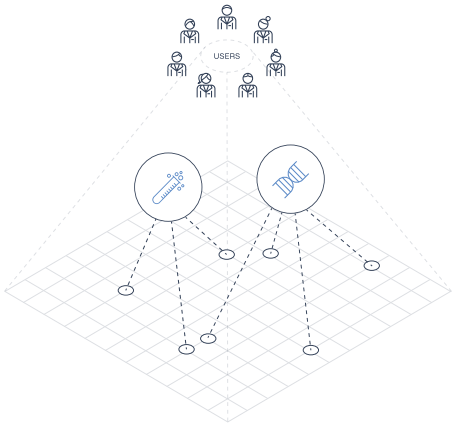
→ The high cost of publishing means it's only ever attempted for a small subset of the research that an institution funds. Scientists have no way of knowing whether the interesting but tangential result they noticed today could be critical to a breakthrough in ten years, and they shouldn't have to. Foundry reduces the cost of sharing final results to nearly zero, allowing researchers to share more of their insight and methods with approved collaborators without slowing down their inquiry.

→ Many studies produce useful intermediate datasets or tools, even if they never get to a conclusive result in the end. Foundry's publishing capabilities also allow researchers to share such assets with approved collaborators alongside rich metadata, making them available to accelerate their colleagues' efforts.

Operational Knowledge Graph

→ Since data access is mediated through Foundry, it is aware of every study the organization conducts. Each study is represented as its own investigation which is automatically linked to relevant real-world concepts (like diseases or compounds) and indexed in the institution's knowledge graph. Since the investigation contains what data was used, what analysis was done, and what results were found, every unit of knowledge is reproducible and verifiable indefinitely into the future.

MANAGING GREAT DATA SCALE AND DIVERSITY



It's no secret that there has been an explosion in data scale in research in recent years. "Big Data" initiatives have dominated the last several years of IT investment in the research space. This focus distracts from an even more complex and nuanced problem: data diversity. Increasingly, we see studies involving clinical data, real-world data (RWD), imaging, -omics, and pre-clinical data. While research institutions already struggle to manage their data, the near future points to an even greater proliferation of scale and diversity. Hospitals and consumers alike are equipped with ever-growing sensor arrays. As of 2020, about one in five Americans use a fitness tracker and since 2001, the cost of sequencing an entire human genome has fallen from $100,000,000 to $1,000.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

This exciting trajectory should also give researchers pause. Growing sensor arrays yield a more complete picture, but without exceptional care they will also yield a more fractured picture. Technical and procedural challenges make it hard to combine these various snapshots into something meaningful, leaving scientists to hop between vendor-specific systems. If scientists are to construct a holistic understanding of human health, their tools must become more powerful, more plentiful, but also more integrated with one another.

04

How Palantir's Foundry Integrates Tools To Build Holistic Understandings →

Native High-Scale Compute, or Bring Your Own Supercomputer

→ Foundry was built from the start to support arbitrary data scale. Foundry's cloud infrastructure allows most use-cases to be scaled up by simply scaling up the servers doing the computation. In places where a High Performance Computing (HPC) is already available, Foundry can connect to it and ensure that you are getting the most out of your investment.

Intergration With Source Systems

→ Foundry's core capabilities as a data integration platform remain best in class. Integrating with a variety of source systems allows researchers to interact directly with everything your organization knows about a problem. All data is traced back to its origin so researchers can be confident in the data they rely on.

Flexible Data Representation

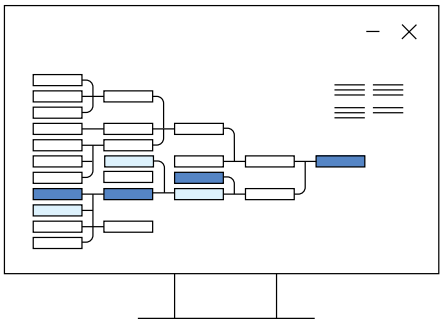
→ The Foundry ontology is highly flexible, allowing new data sources and data types to be represented in-platform and made available to users quickly. As your understanding of the data and its uses evolves, so can its representation in Foundry.

Time-Series, Video, and Image Processing Ahead of Its Time

→ The diverse and pioneering customers who we serve constantly push the platform's capabilities, allowing us to bring technologies from other domains ahead of their need. Foundry has a variety of capabilities we developed for other use-cases that are years ahead of their equivalents in the research space, including time-series analysis perfected in aviation and video/image processing perfected in manufacturing.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

INCREASINGLY SENSITIVE DATA, INCREASINGLY RISKY ENVIRONMENTS



The proliferation of data carries ethical, legal, and financial risk to the modern research enterprise. Each new row and column of data that is integrated, while potentially of scientific value, increases the risk of patient re-identification, medical discrimination, and data leakage. In order to mitigate this, institutions are obliged to limit the scope of data access granted to researchers. In lieu of technologies to enforce such limitations, governance and legal teams often adopt byzantine processes, in some sense relying on security by obscurity. These processes help protect against human error, which can quickly destroy public trust and fetch eye-watering fines, but they also create chilling effects on researcher collaboration.

Data governance must be considered both at the researcher level and at the institutional level. For researchers, a good data protection system gives them confidence - not fear - that their actions are compliant with all relevant policies. For institutions, enforcement technologies must be highly legible so they can understand what data is being used where, by whom, and why. Data protection laws and best practices are also fields actively in flux. An ideal system allows institutions to seize all possible scientific value from their data by utilizing the latest anonymization techniques, while also de-risking the potential for massive overhauls to privacy laws, like GDPR's tumultuous rollout in Europe.

A research institution's intellectual property is among its most valuable assets, which has not gone unnoticed by a variety of adversaries. The line between corporate espionage and cyber warfare grows blurrier by the day. In 2020, Russian and Chinese state-backed hackers attacked vaccine development efforts at multiple pharmaceutical companies. It is clear that highly capable, highly motivated attackers are taking more interest both in source data and the intellectual property stored at research institutions. To survive in this environment, institutions must equip their researchers with the most secure, most legible, most usable tools they can.

05

How Foundry Security Empowers Researchers And Their Institutions →

Deploy Multiple Anonymization Techniques

→ Different scientific use-cases carry different risk profiles and warrant different approaches to anonymization. Foundry enables governance teams to adjust the anonymization technique on a study-by-study basis, and even allows dynamic upgrading and downgrading of data sensitivity within the same study. This allows for basic questions to be answered on less sensitive data tiers, but then upgrading sensitivity as conviction and scientific need grows. In addition to minimizing data exposure, this dynamic anonymization also makes it easy to try new anonymization techniques and stay ahead of the curve.

→ Learn more about protecting patient privacy here [Link]

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

Military-Grade Security, Consumer-Grade Usability

→ Tested and trusted by militaries, intelligence agencies, and secretive companies, Foundry is used by some of the most security-conscious organizations on earth. Full data encryption and multi-layered security perimeters means that the human operator is the weakest link. By providing legible, usable methods for secure research we give end-users the ability to collaborate confidently within the bounds established by their institution.

Programmable Data Governance

→ Foundry has out-of-the-box data governance capabilities that provide governance teams, data owners, and data consumers with clear understanding of data use across the platform. Every customer's threat landscape is unique, so governance workflows can be adjusted to match their risk assessment. We have implemented the full gamut of governance approaches from fully distributed stewardship, where no single user can grant access to any data, all the way to centralized stewardship where a single committee controls data from multiple institutions, and everything in between.

Source Data and Derived Knowledge, Secure Together and Apart

→ Foundry manages both source data and the knowledge derived from that data, putting both assets behind a military-grade security perimeter and linking each finding to the study and data that produced it. Despite that linkage, Foundry provides granular control over who gets access to a study's source data and separately who gets access to the findings from that data. This allows organizations to choose whether they want all derivative work to be secured by the policy set on the original data, or they can facilitate discoverability of scientific results and metadata without exposing source data.

LOOKING FORWARD

While Palantir cannot predict the future, we can extrapolate from a pool of deep partnerships across the research space. What we see is that just as individual people and teams are less equipped to make serious scientific progress, so too are individual institutions. In aggregate, the scientific community has more capacity than ever before. Since that capacity is split across institutions who struggle to collaborate, however, it's impossible to coordinate it en masse against scientific problems.

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.

Barriers to collaboration, which are technical, legal, and incentive in nature can be solved by technology. The key requirements of said technology are robust security for data, logic, and results; a diverse set of tools targeted at diverse skillsets; and provenance to ensure that collaborators receive proper credit for their work. Foundry has already been used to facilitate several pioneering cross-institutional collaborations, with extremely promising results, outlined below.

As the community's untapped capacity grows and its unsolved problems become more challenging, multi-institutional collaboration will become an existential imperative. We believe that the organizations who proactively address the challenges above will be situated for success in this new environment, and the ones who do not will face acute pressure to adapt.

The frontier of knowledge is the boundary between the solved and unsolved mysteries of our universe. The easy problems were answered yesterday. For tomorrow, only the most difficult ones remain. What must we do today? Let's explore how Foundry solves these challenges today and positions research institutions for the opportunities of tomorrow.

REAL WORLD IMPACT

One exciting data point for the value of multi-institutional scientific research emerged through our partnership with the National Institutes of Health (NIH). The NIH has been a long-time partner and, as such, has been on the forefront of many innovations mentioned above.

In response to COVID, NIH was able to boot up the National COVID Cohort Collaborative (N3C) on top of Palantir Foundry in a matter of weeks. Today, there are over 9 billion rows of data, more than 250 research projects, and participants from more than 70 institutions (CD2H), forming the world's largest centralized patient-level data asset of COVID-19 clinical data. There is no shortage of data and expertise in the N3C enclave, and scientists have taken the initiative to create useful tools for one another.

A single scientific artifact, made reusable by Foundry, was used by more than fifty research projects on which members of multiple institutions have worked. We estimate this saved upwards of 10 aggregated years of researcher time in a period where faster cycle times directly correlated to lives saved.

Learn More → Impact

Copyright © 2021 Palantir Technologies Inc. and affiliates ("Palantir"). All rights reserved. The information in this document is proprietary and provided for informational purposes only and shall not create a warranty of any kind. Any data contained herein is notional.