whitehat analytics

Data science
transforming the pharmaceutical landscape

About the authors

Katrin Wiederhold, PhD, is Director of Life Sciences at Whitehat Analytics and has worked in Life Sciences for over 15 years. Since her education at the Max Planck Institute and MRC Laboratory of Molecular Biology (Cambridge) she has been working in bleeding edge immunotherapy companies driving their vision, objectives, organisational structure, culture & values. At Whitehat Analytics Katrin is ensuring that our Life Science customers' challenges are efficiently addressed by contributing her deep understanding of the sector.

Tobias Kloepper, PhD, founded his first business while studying for his Diploma in Mathematics and Computer Science. Putting this grounding to good use, he then completed a PhD in Biological Informatics. His enviable scientific alumni include the MRC Laboratory of Molecular Biology (Cambridge), the University of Oxford and the Max Planck Society. Tobias founded Whitehat Analytics in 2015 to help businesses reinvent themselves using empirical data-driven strategies supported by artificial intelligence.

HOW DATA SCIENCE CAN TRANSFORM THE PHARMACEUTICAL LANDSCAPE

For every 5,000 compounds starting in the laboratory, five are tested in humans and one makes it to market. Moreover, it takes approximately 10 years and an average cost of $2-3 billion to develop each new drug.[i] During this development, a vast amount of molecular and clinical data are created and subsequently stored in proprietary networks.

Significant effort is currently being focused on data management at the start of the drug development process, in the drug discovery phase. However, the substantial impact data science can have further downstream is largely neglected, meaning there is a huge potential opportunity for pharmaceutical businesses to digitally transform their downstream operations.

In this whitepaper we outline how data science can have a transformational impact on big pharma, particularly in the drug manufacturing and commercialisation stages.

1. WHERE ARE BIG PHARMA ORGANISATIONS CURRENTLY INVESTING THEIR MONEY?

With disease complexity more apparent as research technology becomes more sophisticated, developing new therapies has never been more complex, or more costly. R&D spend in the UK pharmaceutical sector alone is at an eight-year high, rising to £4.8bn in 2020, an increase of £306m on 2019 and representing the largest real-term increase of any sector.[ii]

The main focus of large pharmaceutical organisations is on the clinical development and commercialisation of drug candidates. This includes drug development costs such as manufacturing, clinical trial costs and the whole process of commercialisation including production and marketing. A study of the top 10 US-based drug makers by CSRxP and GlobalData found that, while drug companies spend identical revenue percentages on R&D and on operation and production (22% each), they also spend nearly one fifth (19%) – or $47 billion – on marketing, advertising and promotion.[iii]

Extensive use of data science within drug discovery at the clinical trial stage is well established, or is at least under intense investment. The real – and as yet untapped – opportunity for drug developers to implement an effective data science program lies within manufacturing and commercialisation

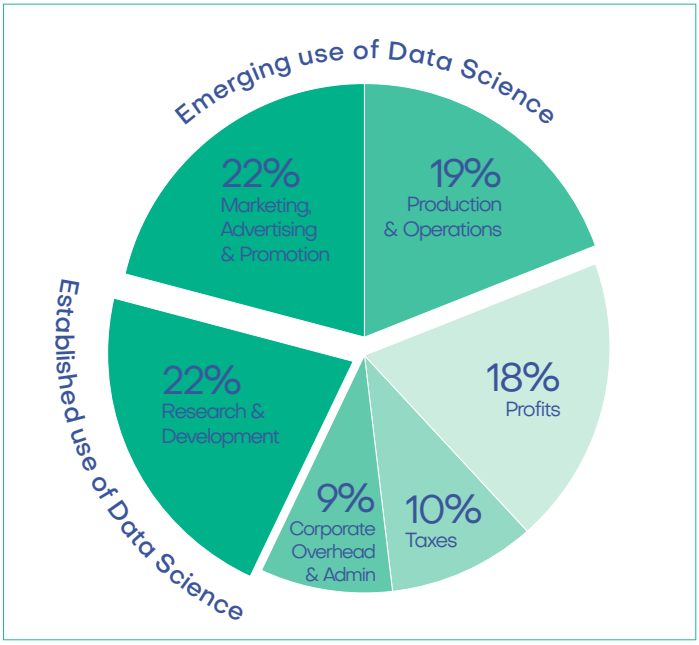

Figure 1: Top pharma percentage spend
While data science projects are well established within research and development this reflects only 22% of the total spend. Moving data science downstream into the areas of production, marketing and advertising offers the opportunity to streamline a significantly larger part of the budget. We consider this the next big opportunity in the market and see initial projects emerging into production.

2. REALISING THE OPPORTUNITY FOR DATA SCIENCE IN THE PHARMA SPACE

Bringing a new pharmaceutical product to market is a lengthy process with many bottlenecks, diversions and U-turns along the way. Trials regularly fail to meet their initial objectives which can add significant delay and increase the costs of an already expensive investment. The vast amount of data generated throughout the drug development process makes implementing a cohesive data science strategy at every stage the next logical step.

The application of cutting-edge data technologies, machine learning and continuous deployment infrastructures are starting to bring us closer to matching the dynamic nature of the drug development process. This means we may finally be able to translate digital initiatives into the long-promised operational performance improvements, ultimately bringing more effective drugs to market faster and more cost-effectively.

Many of the big players in the pharmaceutical industry have started major initiatives for the application of these technological advancements in the drug discovery and clinical phases. These have proven to be able to extract hidden insights from collected data and allow for more efficient execution of drug development

Novartis case study: a model of good practice[iv]

Forward-thinking pharmaceutical manufacturers are already reaping the benefits of investing in data science.

The Novartis "Go Big on Data" initiative[iv] is starting to fundamentally change how the pharmaceutical giant operates. Using Amazon Web Services' (AWS) portfolio of artificial intelligence (AI) and machine learning (ML) cloud services, Novartis built an enterprise-wide data and analytics platform that will transform its business operations and change the way that its medicines are manufactured and delivered. The project will put real-time analytics in the hands of associates, helping drive agility, innovation, and cost efficiencies across Novartis' global business processes and systems, empowering them to make better decisions and increasing efficiencies throughout the supply chain.

The new platform

- ingests data from disparate sources, cleans it, links it, and makes it ready for analytics to be applied
- enables sophisticated ML algorithms to be built that can analyse the data and generate actionable insight
- enables web-based solutions/applications for employees to use in their workflows to improve their decision making.

In setting out to access experience, create intelligence and unlock value, Novartis estimates it has achieved productivity gains of around 10% across its portfolios since the application of the new data science platform.

3. LEVERAGING AGILE DATA

The drug commercialisation process is one of the most challenging steps in drug development and the best way to create a successful strategy is to take guidance provided by reliable data from multiple sources. In an increasingly competitive landscape, pharma companies need powerful solutions – and that means leveraging the power of cutting-edge technology.

The marketing and promotion of a drug is another area where we currently see significant opportunity for data science. With increasing competition from generics, big pharma is getting smarter about analysing and driving effectiveness in its sales and marketing operations – evidenced by the high spend. This is a highly regulated environment, in which particular care and attention must be taken with innovative concepts and claims must be supported by data. Bringing the concept of Agile data science into the regulated environment can deliver a powerful new competitive advantage.

New, niche and under-served markets may be spotted by combining and analysing information from social media, demographics, electronic medical records and other data sources. Equally, analysing the effectiveness of sales efforts and capturing the feedback and using it effectively can help pharmaceutical companies get an edge on their competition.

4. THE ROLE OF DATA SCIENCE IN PHARMA

Pharmaceutical companies have long relied on empirical data to drive research, identify patterns, test theories and understand the efficacy of treatments. Data models can be applied throughout the drug development process to establish patterns of best practice and create predictive models for future success or failure. Here we explore how data science can be applied to the discovery, manufacturing, clinical trial and commercialisation phases of drug development.

4.1 Discovery

Big data and machine learning are increasingly applied to the discovery phase of drug discovery. For example, applying data science to cell therapy R&D can reduce risks and accelerate development. Single-cell next-generation sequencing data is uncovering several novel cell types, many of which are appealing for cell therapy.

Data-driven tools can also provide predictions of perturbations to direct cell fate systematically. For instance, it is now possible to use data-driven tools to predict which transcription factors are required to convert between any two human cell types. Combined, these will enable a new data-driven development pipeline for cell therapies.

Data science also gives researchers the ability to generate ideas for entirely novel compounds, where the 'invented' molecule is predicted to have all the desired properties required for success – which could hugely accelerate the discovery of effective new drugs. For example, an AI network developed by Google AI offshoot DeepMind has made a significant step in solving one of biology's greatest challenges – determining a protein's 3D shape from its amino acid sequence. The ability to accurately predict protein structures from their amino acid sequence would be a huge boon to medicine by vastly accelerating efforts to understand the building blocks of cells and enable quicker and more advanced drug discovery.

4.2 Manufacturing

Pharmaceutical manufacturers are increasingly shifting from blockbuster drugs to biological and genomic medicines that have shorter life cycles. And with biological medicine accounting for more than a quarter of the entire pharmaceutical market,[vi] manufacturers must be able to deliver these sensitive medicines within tighter time frames.

Manufacturing of pharmaceuticals follows validated processes to ensure consistency and quality of the final drug product. Data science can help to support quality control in many ways, for instance by, for example, providing automated and reproducible analytical pipelines that utilise predictive models to identify adverse events. With the power of in-memory computing technology and interconnected and automated systems, manufacturers can analyse large amounts of quality, environmental and

Internet of Things (IoT)-generated data. Tapping into this Big Data allows pharma companies to build AI-driven end-to-end process controls, resulting in higher quality products, more predictability, more efficient manufacturing and faster time to market.

4.3 Clinical trials

There is considerable pressure for life science companies to find ways to fail proposed products faster and to identify successful new products more quickly, to expedite launch to market. Integrating clinical data enables the use of machine learning to speed up the analysis of the data. Often this can lead to real time insights into the clinical trials and support critical decision making. This is proving to provide efficiency gains as well as improve safety of clinical trials.

Clinical trials are costly and time consuming to run and pharmaceutical companies must ensure that they have recruited the right mix of patients for a given trial. Big data can assist in identifying the appropriate patients to participate in a trial through analysis of demographic and historical clinical data, remote patient monitoring, reviewing previous clinical trial events, and helping to identify potential side effects before they become a reality. Patient data will help pharmaceutical companies consider more factors, such as genetic information, to help companies identify niche patient populations in sources such as social media, which in turn will speed up and reduce costs of trials.

Using this approach, clinicians can understand various medical and lifestyle details of every patient and analyse their suitability for a clinical trial.

Big data can also assist by aggregating data across multiple clinical trials and look for patterns that are beyond the specific use case to create predictable foresights. Without this, performing cross-trial analysis is a daunting and costly task that requires programmers and developers to integrate data manually, which is a time-consuming endeavor that is prone to human error.

It is projected that implementing rudimentary machine learning and data science into the clinical trials process alone can and has been cutting costs up to 20%.[vii]

4.4 Commercialisation

AI and machine learning can deliver previously inaccessible insights that positively impact the commercialisation aspect of drug development. They can inform how companies deploy precision strategies for the best possible returns with the right balance of resources, speed and at scale.

AI and machine learning methods consistently deliver more accurate outcomes in less time than conventional assessments. In the short term, that translates into competitive advantage and better sales results. Over time these automated solutions deliver extended value by continuously monitoring trends and optimising results as new data is generated.

There is huge potential to make better use of the vast amount of data across a wide range of areas affecting drug commercialisation, including:

- Understanding market dynamics, competitive strategies, clinical practices, regulatory issues, access challenges and lifecycle planning
- Establishing strategies for pricing, market access and reimbursement in multiple geographies
- Developing strategies for internal and external stakeholders who influence the commercialization of the brand
- Supporting brand strategy planning, brand forecasting, strategic clinical development options and communications strategies.

5. ESTABLISHING STRONG DATA GOVERNANCE

The primary goal of data governance is to ensure that data assets meet an enterprise's standards in terms of its integrity, compliance with regulations and quality. Having strong data governance processes and structures in place provides not only regulatory safety, but more importantly enables the use of data in clearly defined boundaries, ensuring that a wide range of use cases can be applied. Strong governance sets the foundation for the sustained commercial success of data science. For example, the use of clinical trial data beyond its primary indication is tightly regulated. It can often be the case that regulatory frameworks exist within a company, but across multiple trials can vary significantly, which has a negative impact on the reuse of these data repositories.

Once a coherent governance framework is in place, the FAIR (findable, accessible, interoperable and reusable) data principles should be applied (Figure 2). Binding FAIR principles into governance frameworks can advance the digital transformation of the pharma industry, enabling pharma businesses to leverage a plethora of operational efficiencies while also reducing time to market and cutting development costs.

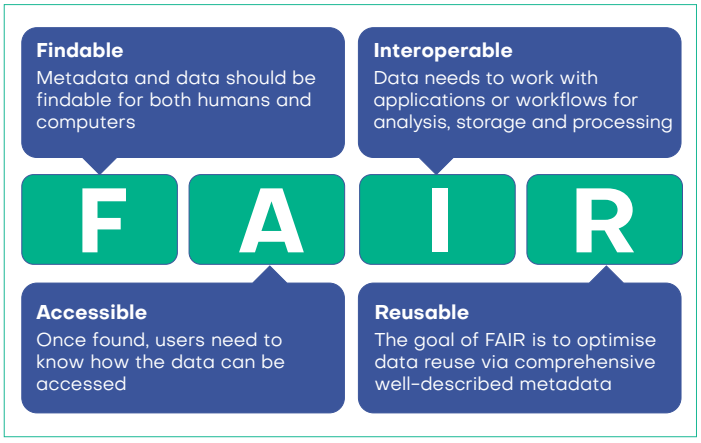

Figure 2: FAIR Data Principles

There are several challenges that pharma organisations are currently facing in establishing a universal governance framework. Date may be siloed across departments – and large amounts of new data are continuously being created. Another challenge is to keep governance teams up to date with the exceptionally fast developing data science field, particularly on aspects that are of high business interest to the pharma industry. As a consequence, pharmaceutical companies are often not in a position to make strategic decisions based on their data and are even less well placed to create efficient data science functions.

These technology integration challenges hamper all areas of drug development - from sales and marketing to research, clinical and manufacturing. Service oriented architecture (SOA) is emerging as the new standard for successful IT integration, to help address these challenges across the organisation and serve as the blueprint to align business R&D with data science. [viii]

6. HOW CAN SOA HELP GOOD GOVERNANCE?

SOA governance refers to the processes used to oversee and control the adoption and implementation of SOA in accordance with recognised practices, principles and government regulations. It helps optimise service quality, consistency, predictability and performance, ensures that personnel follow prescribed policies and corrects system problems or policy infractions as they occur.

Before SOA emerged in the late 1990s, connecting an application to data or functionality housed in another system required complex point-to-point integration - integration that developers had to recreate, in part or whole, for each new development project. Exposing those functions through SOA eliminates the need to recreate the deep integration every time. All data is put into a storage environment which lets individuals pick what they want from the data through using data indexing, a structure technique that is used to quickly locate and access the data in a database. It collects, parses and stores data to enhance the speed and performance of retrieving and analysing relevant documents.

The collaborative nature of SOA can help address a new level of partnering and global reach that is necessary to stay competitive in the pharma industry both for today and the future.

7. BENEFITS OF A DATA-DRIVEN CULTURE

Data science is a collaborative effort; all stakeholders must be organically integrated and work for the same cause. A data-driven culture needs to be accepted across the organisation as it will be beneficial for all departments.

Pharma organisations have the new opportunity to develop a cultural framework that helps all members of the organisation to move data to the centre of decision-making. Open access to data reinforces a culture of democracy, as every department and every individual is aware of the big picture and the role they play within the organisation.

One challenge for big pharma is that the stakeholders have been working in siloes for many decades, and the territorial mindset pervades. For data science to make an impact, stakeholders must sit at the decision-making table rather than just being service providers.

Culturally, senior leadership buy in is often the key factor. Their willingness to push a data-driven culture throughout the organisation will engage every individual and ensure that everyone is working towards the same goal – and sees it as a critical part of their own development and reward.

8. THE CONVERGENCE OF SCIENCE AND DATA

It is clear that the diverse fields of life sciences, computer science and data science are converging. In 20 years, being data-driven will be the norm; forced by economics, particularly in the pharma space where efficiency gains are significant.

The future success of pharma organisations will depend largely on an ability to bring data to the forefront of business decision making. Although technical and cultural challenges undoubtedly lie ahead, advanced data analytics represent the biggest growth opportunity that pharma companies should look to leverage.

References

[i] Master's in Data Science, Data Science in the Pharmaceutical Industry, June 2020. https://www.mastersindatascience.org/industry/pharmaceuticals/

[ii] UK pharma sector R&D spending at 8-year high, Drug Discovery World, November 2020. https://www.ddw-online.com/uk-pharma-sector-rd-spending-at-8-year-high-8465-202011/#:~:text=Total%20pharmaceutical%20industry%20spending%20on,sector%2C%20according%20to%20the%20data.

[iii] Dose of reality breaking down the pharma dollar, The Campaign for Sustainable Rx Pricing, 2020. https://www.csrxp.org/dose-of-reality-breaking-down-the-big-pharma-dollar-part-iii/

[iv] Amazon Web Services announces strategic collaboration with Novartis to accelerate digital transformation, Novartis.com, December 2019. https://www.novartis.com/news/amazon-web-services-aws-announces-strategic-collaboration-novartis-accelerate-digital-transformation-its-business-operations

[v] Digital transformation: a Novartis review, D4 Pharma, July 2020. https://d4-pharma.com/digital-transformation-a-novartis-mini-review/

[vi] Enhancing the pharmaceutical supply chain with big data, Pharma Manufacturing, January 2020. https://www.pharmamanufacturing.com/articles/2020/enhancing-the-pharmaceutical-supply-chain-with-big-data/

[vii] Krigsman (2020) Data Science in Pharma and Drug Discovery. https://www.cxotalk.com/episode/data-science-pharma-drug-discovery#_Toc31202659

[viii] Blueprint for successful SOA integration, Oracle, 2008. https://www.oracle.com/technical-resources/articles/enterprise-architecture/blueprint-soa-integration.html

Note: Whitehat Analytics has not been involved with the Novartis digital project. It is included as a model of good practice.

About Whitehat Analytics

We enable companies to be powered by their data.

To prepare themselves for the information age, companies with legacy technologies, processes and skillsets must undertake a business transformation to become data-led organisations.

Our mission is to help companies undertake that journey.

Our goal is to help companies to structure themselves around an empirical, data-driven business model, where decisions and process are analytical and outcome-focused.

By developing a data-led culture that runs through the entire organisation, we allow our clients to unlock the power of data and thrive in a more competitive, data-led global marketplace.

We drive cultural change: we embed analytical capability deep within the organization, via a structured program of engagement, education, and enablement. All parts of the organization are empowered to use data to do their jobs better.

We deliver technical excellence: we work on the bleeding edge of data science, integrating large, complex data sets with the latest tools, and using the most advanced analytical techniques such as neural networks to allow our customers to see further and react faster.

We are compliance-driven: we know your data must be secure, it must be accounted for, and the quality must be verified. Your businesses reputation relies on your customers trusting you and your processes, so we build in governance, audit, compliance, and quality validation at every stage of the process. We are ISO 9001 and ISO 27001 certified – so we understand the compliance process first-hand.

We are outcome-focused: we believe that data science must be deployed to drive business value. That's why we ensure that we deliver products and services that are usable, scalable and robust, to change business outcomes on the frontline.

whitehat analytics

20-22 Wenlock Road
London
N1 7GU
United Kingdom

+44 (0) 203 603 2773
contact@whitehatanalytics.com
www.whitehatanalytics.co.uk